



Organisation de la chromatine et signalisation par les oestrogènes

Justine Quintin

► To cite this version:

Justine Quintin. Organisation de la chromatine et signalisation par les oestrogènes. Médecine humaine et pathologie. Université de Rennes; Université européenne de Bretagne (2007-2016), 2013. Français. NNT : 2013REN1S074 . tel-01135877

HAL Id: tel-01135877

<https://theses.hal.science/tel-01135877>

Submitted on 26 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention : Biologie

Ecole doctorale Vie Agro Santé

présentée par

Justine Quintin

Préparée à l'unité de recherche UMR CNRS 6290
Institut de Génétique et Développement de Rennes
UFR Sciences de la vie et de l'environnement
Université de Rennes 1

**Organisation de la
chromatine et
signalisation par les
œstrogènes**

**Thèse soutenue à Rennes
le 6 mars 2013**

devant le jury composé de :

Kerstin BYSTRICKY

Professeur d'Université, UMR CNRS 5099 –
Université Paul Sabatier-Toulouse / *rapporteur*

Cécile ROCHETTE-EGLY

Directeur de Recherche CNRS, IGMBC –
Strasbourg / *rapporteur*

Christian JAULIN

Directeur de Recherche CNRS, IGDR - Université
de Rennes 1 / *membre*

Gilles SALBERT

Professeur d'Université, UMR CNRS 6026 –
Université de Rennes 1 / *directeur de thèse*

RESUME :

En réponse à son environnement composé de signaux endogènes et exogènes, une cellule doit pouvoir adapter son transcriptome et cela à travers une modulation fine de l'expression de ses gènes. Les mécanismes permettant une telle adaptation reposent sur de multiples paramètres, entre autre l'organisation du génome, que ce soit au niveau de sa séquence primaire ou de son organisation au sein de la chromatine qui est un support pour l'intégration de nombreuses informations (structurelles et épigénétiques). De plus, l'organisation tridimensionnelle du noyau cellulaire apporte des contraintes physiques et fonctionnelles qui contribuent également à ces régulations. Afin de comprendre comment toutes ces informations peuvent être intégrées lorsqu'un signal régule la transcription d'un ensemble de gènes colinéaires ("cluster" de gènes), nos études se sont focalisées sur la description et dissection des mécanismes impliqués dans la régulation coordonnée de gènes œstrogéno-dépendants par le récepteur aux œstrogènes (ER) dans des cellules cancéreuses d'origine mammaire. Dans ce cadre, nous nous sommes plus particulièrement intéressés au cluster TFF, situé sur le bras long du chromosome 21, incluant le gène modèle *TFF1*, en utilisant des techniques d'analyse à grande échelle (ChIP-chip, ChIP-seq, 4C et analyses transcriptomiques).

La première partie de ce travail a consisté en (i) l'étude de ces mécanismes en intégrant les aspects régulateurs à la fois dans l'espace et le temps; ainsi qu'en (ii) l'identification des facteurs protéiques collaborant avec ER pour la mise en place de ces processus. Cette étude, menée dans deux types cellulaires différents (les MCF-7 qui expriment ER de façon endogène et MDA::ER qui l'expriment de façon ectopique) a permis de mettre en évidence un réseau d'interactions fonctionnelles cellule-spécifiques mettant en jeu des contacts dynamiques entre sites de liaison d'ER distants et promoteurs de ces gènes cibles, ainsi qu'un rôle des protéines de la famille GATA dans l'établissement et/ou la fonctionnalité de ce réseau.

La seconde partie de ce travail de thèse a porté plus particulièrement sur une autre famille de protéines impliquées dans l'établissement du cistrome ER: les facteurs FOX. En effet, la ré-expression de ER dans les cellules MDA::ER ne reproduit ni un cistrome ER ni un transcriptome œstrogéno-dépendant ressemblant à ceux des cellules MCF-7. Nous avons posé l'hypothèse que ceci pouvait être dû à l'absence dans les MDA::ER du facteur pionnier FOXA1. Après la création d'une lignée cellulaire MDA::ER exprimant FOXA1 et les analyses subséquentes, nous avons pu caractériser des mécanismes d'interrelations interférentes entre le FOXA1 ectopique et du FOXA2 endogène qui, pourraient être à l'origine du fait que FOXA1 soit incapable de réorienter les cellules MDA::ER vers un phénotype prolifératif œstrogéno-dépendant ressemblant à celui des cellules MCF-7.

Remerciements

Je remercie Claude Prigent, directeur de l'UMR 6290, de m'avoir accueillie mais également Gilles Salbert, directeur de l'UMR 6026 et chef de l'équipe SPARTE au sein de laquelle j'ai effectué mes 3 ans et demi de thèse.

Je tiens à remercier particulièrement Kerstin Bystricky et Cécile Rochette-Egly d'avoir accepté de juger mon travail de thèse en qualité de rapporteur malgré un emploi du temps chargé. Je souhaite également remercier Christian Jaulin d'avoir accepté de faire partie de ce jury de thèse.

Je souhaite remercier en premier Luc Paillard et Vincent Legagneux, car ce sont vous qui êtes à l'origine de ma présence dans l'équipe SPARTE et finalement à l'origine de beaucoup de chamboulements dans ma vie, entre autre une absence de sommeil depuis le 17 Décembre 2012, rien que pour ça, je pense que si un jour je suis en galère de nounou, vous êtes obligés de me dépanner ...

Un grand merci à tous les membres de mon équipe, Gilles de m'avoir acceptée dans son labo et d'avoir été, je pense, le meilleur chef d'équipe qu'on puisse souhaiter. Maud, Christine, GP et Cath1 pour la bonne humeur qu'elles mettent tous les jours, pour nos restos filles toujours synonyme de « grandes révélations », pour les coups de mains, les absences de coup de gueule... Christine pour les bons conseils, qu'ils soient perso ou professionnels, et pour les corrections de cette thèse sauf de ces remerciements qui du coup doivent être truffés de fautes, oups ! Maud pour les interrogatoires des diverses personnes pour savoir qui est enceinte, pour les aides sur les manips, sur les proto sur tout finalement, et GP pour ses fabuleux karaokés dans le labo, en particulier sur « Sexy and I know it » ☺ et les cancans, merci à Elise, bonne chance pour tes années de thèse, profite en ça passe vite, très vite et enfin merci à GDD, quels 6 mois en ta présence ! Grosses larmes, fous rires, c'était un plaisir de bosser avec toi.

Je tenais également à remercier l'équipe HIP et en particulier Pascale qui voit toujours tout du bon côté, tout est beau, tout le monde est gentil. Florence pour ses petites piques toujours très drôles et ces histoires sur sa petite deuxième, au moins je sais à quoi m'en tenir pour le futur.

Mention spéciale pour Rhodiannia pour tellement de choses que je ne sais même pas quoi citer, les anticorps ? le crayon Mickey ? l'escapade à Disneyland ? ou peut être pour m'avoir permis de voir GP sortir de ses gonds !

Fred P, pour les conseils, les blagues, l'aide pour le déménagement, les reprises des tubes des années ?... je ne sais même pas tellement les chansons sont vieilles, les choré ... Juste une déception, ne pas avoir fait au moins une fois une soirée mousse dans cette incroyable salle de culture, tant pis.

Vous êtes tous, je pense, en grande partie responsable de ma difficulté à arrêter la biologie ou plutôt à quitter SPARTE. Bien évidemment les manip vont me manquer, après 8 ans passer avec une pipette à la main ce n'est pas simple, mais c'est l'ambiance fabuleuse de l'équipe et la super entente dans le bâtiment qui vont être le plus dur à quitter même si malheureusement la biologie n'est pas faite pour moi ou plutôt je ne suis pas faite pour la biologie !

Bien sur, merci à tous les membres des équipes HIP et RED ainsi qu'à Colette, Elizabeth, Mado, Charlotte, Ariana (félicitation et bonne chance aux states),Yohann, pour votre gentillesse, sympathie, disponibilité...

Mention spéciale à Raphael qui m'a supporté pendant 6 mois de M2 puis 3 ans de thèse et surtout pendant les 9 derniers mois qui ont été particulièrement fatigants, euh en fait non c'est l'issue de ces 9 derniers mois qui est fatigante mais ça faut pas le dire, c'est une surprise pour Maud. Bref, MERCI.

Et bien sur, ne vous inquiétez pas, vous entendrez régulièrement ma douce voix dans les couloirs... d'ailleurs je crois que je dois 5 pots, arghh !

ABBREVIATIONS :

3C: "chromosome capture conformation"

5-hmC: "5-hydroxymethylcytosine"

5-mC: "5-methylcytosine"

A.A: acide aminé

ADN : acide desoxyribonucléique

ADP : adenosine diphosphate

AF : "activation function"

AID : "activation induced deaminase"

APOBEC : "apolipoprotein B mRNA editing enzyme"

AR : "androgen receptor"

ARN: acide ribonucléique

ARNi: ARN interference

ARNm: ARN messenger

ARNs nc: ARNs non codant

ARN-PII : ARN polymerase II

ARNr: ARN ribosomique

ARNt: ARN de transfert

ATM : "ataxia telangiectasia mutated"

ATP : adenosine triphosphate

BAFs : "Brg1 associated factors"

BER : "base excision repair"

Brg1 : "Brm-related gene 1 protein"

hBrm: Brahma

CBP : "creb binding protein"

CHD : "chromodomain helicase DNA binding protein"

ChEC : "chromatin endogenous cleavage"

ChIP: "chromatin immunoprecipitation"

COUP-TF: "chicken ovalbumin upstream transcription factor"

CPI: complexe de pré initiation

CPSF: "cleavage polyadenylation specificity factor"

CREB: "cAMP response element"

CstF: "cleavage stimulation factor"

CTs: "chromosome territories"

CTD: "C terminal domain"

CTE: "C terminal extension"

DBD: "DNA binding domain"

DNMT: "DNA methyl transferase"

ER: "estrogen receptor"

ERE: "estrogen responsive element"

ERR: "ER-related receptor"

ES: "embryonic stem cell"

E1 : œstrone

E2 : 17 β -œstradiol

E3 : œstriol

FAIRE : "Formaldehyde Assisted Isolation of Regulatory Elements"

FISH : "fluorescent *in situ* hybridization"

GFP : "green fluorescent protein"

GO: "gene ontology"

GR: "glucocorticoid receptor"

GTFs: "general transcription factors"

HAT: "histone acetyl transferase"

HCP: "High CpG promoter"

HDAC: "histone deacetylase"

HMT: "histone methyl transferase"

HSP: "heat shock protein"

HP1: "heterochromatin protein 1"

ICD: "intra chromosomal domain"

ICN: "intra chromosomal network"

ICR: "imprinting control region"

KO: "knock out"

KRAB: Krueppel associated box

LAP2 β : lamina-associated polypeptide 2 β

LBD: "ligand binding domain"

lincARN: "long intergenic non coding RNA"

LINEs: "long interspersed nuclear elements"

MAPK: "mitogen activated protein kinase"

MBD: "methyl binding domain"

MeDIP: "methylated DNA immuniprecipitation"

miRNA: "micro RNA"

MR: "mineralocorticoid receptor"

NORs: "nucleolar organizer regions"

PR: "progesterone receptor"

NPC : "nuclear pore complex"

Nups : nucléoporines

OPT : "Oct1/PTF/Transcription"

Pb: paire de base

PBX1: "pre B cell leukemia homeobox-1"

PcG: "polycomb group"

PML: "promyelocytic leukemia"

PPAR: "peroxisome proliferator activated receptor"

PRE: "PcG responsive element"

qPCR: PCR en temps réel

RAL : raloxifène

RAR : "retinoic acid receptor"

RXR : "9-cis retinoid acid retinoid X receptor"

SAGA : "Spt-Ada-Gcn5-Acetyltransferase"

SANT domain: "Swi3, Ada2, N-CoR, TFIIB domain"

SINEs: "short interspersed nuclear elements"

siRNA: "small interference RNA"

SLIDE: "SANT like ISWI domain"

snRNP: "small nuclear ribonucleoproteins"
SWI/SNF: "switch/sucrose non fermentable"
SWR1: "swi/snf related protein"
TAFs: "TBP associated factors"
TALE: "three amino acid loop extension"
TAM: tamoxifène
TBP: "TATA binding protein"
TDG: "thymine DNA glycosylase"
TES: "transcriptional end site"
TET: "ten eleven translocation"
TF: facteur de transcription
THC: tetrahydrochrysène
TR: "thyroid hormone receptor"
TSS: "transcription start site"
TRD: "transcriptional repression domain"
VDR: "vitamin D receptor"
Xi: "X inactive"

SOMMAIRE

SOMMAIRE :

SOMMAIRE	1
INTRODUCTION	5
I-Architecture du noyau eucaryote	5
A-L'enveloppe nucléaire	5
1-La lamina	6
2-Les pores nucléaires	6
B-Les corps nucléaires	8
II- La chromatine	10
A-Organisation structurale de la chromatine	10
1-Le collier de perles	10
2-Les histones	10
3-Les histones de liaison	11
4-Les variants d'histones	11
a-Les variants de H2A	11
b-Les variants de H3	13
B-Modifications de la chromatine	14
1-Les modifications des histones	14
2-Les modifications de l'ADN	16
a-Méthylation des cytosines	16
Présentation	16
Mise en place et maintien de la méthylation	17
Rôles de la méthylation	18
b-Hydroxyméthylation des cytosines	19
c-Dynamique des modifications de l'ADN	19
C-Organisation fonctionnelle et condensation de la chromatine	20
1-L'hétérochromatine	20
a-Mécanismes généraux d'hétérochromatinisation	21
b-L'hétérochromatine facultative	21
c-Cas du Xi	22
d-L'hétérochromatine constitutive	23
2-L'euchromatine	24
3-Les grands domaines de chromatine	25

D-Niveaux supérieurs de l'organisation chromatinienne	28
1-Notion de territoires chromosomiques	28
2-Protéines impliquées dans l'organisation des boucles de chromatine	29
a-La cohésine	29
b-CTCF	31
c-Complexes cohésine et CTCF : actions coordonnées et distinctes	32
3-Etudes à grandes échelles des interactions inter et intra chromosomiques (3C, ChIA PET, HiC)	33
III-La transcription	37
A-Généralités	37
B-L'accessibilité de l'ADN	37
1-Les complexes de remodelage ATP-dépendants	38
a-La famille SWI/SNF	38
b-La famille ISWI	39
c-La famille Mi2-NuRD	39
d-INO80	40
e-SWR1	40
2-Les enzymes de modification des histones et les variants d'histones	41
C-Activation de la transcription de gènes ARN-PII-dépendants	41
1-Généralités : promoteurs et séquences régulatrices	41
2-Mise en place du complexe de pré-initiation de la transcription	42
a-Cas des promoteurs TATA+	42
b-Cas des promoteurs TATA-	43
c-Initiation <i>via</i> le complexe SAGA	44
3-Modulation de l'initiation par les promoteurs proximaux	44
a-Assemblages différentiels	44
b-Le complexe « mediator »	45
4-Elongation et terminaison de la transcription	46
a-Synthèse de l'ARNm	46
b-Terminaison	47
5-La dynamique de la transcription	48
a-Au niveau des protéines	48
b-Au niveau des promoteurs	49
D-La transcription aux enhancers	50
E-La répression transcriptionnelle	51

1-La répression par les protéines polycomb	51
2-La répression par p53	52
F-Remarques générales sur l'activation et la répression transcriptionnelle	53
IV- Les récepteurs nucléaires	54
A-La famille des récepteurs nucléaires	54
1-Classification	54
2-Structure	55
3-Mécanisme d'action	56
a-La localisation	56
b-La liaison du ligand	57
c-Dimérisation et liaison à l'ADN	57
d-Activation des gènes par les RNs	58
e-Inhibition des gènes par les RNs	58
f-La voie non génomique	59
B-Le cas du récepteur aux œstrogènes : ER	59
1-Les œstrogènes	59
2-Les récepteurs aux œstrogènes	60
a-ER α	60
b-ER β	60
3-Rôle de ER après liaison d'un agoniste	60
4-Rôle de ER après liaison d'un antagoniste	61
5-ER et la voie non génomique	61
V-Les facteurs pionniers	63
A-Définition	63
B-Les facteurs FOX	63
a-FOXA1	64
b-FOXA2	65
C-Les facteurs GATA	66
D-PBX1	67

RESULTATS	68
I- Partie 1	68
Contexte de l'étude	68
Objectifs	68
Contribution aux différents résultats	69
Publication	70
" Cell-specific and spatio-temporal controls of the estrogen-responsive trefoil factor (TFF) locus activity "	
II- Partie 2	
Contexte de l'étude	141
Objectifs de l'étude	141
Contribution aux différents résultats	141
Publication	142
" A FOXA1/FOXA2 competition may hinder the concerted reorientation of an ER negative cell line towards an estrogen-sensitive growth phenotype by FOXA1 and ER "	
DISCUSSION GENERALE	181
BIBLIOGRAPHIE	188

INTRODUCTION

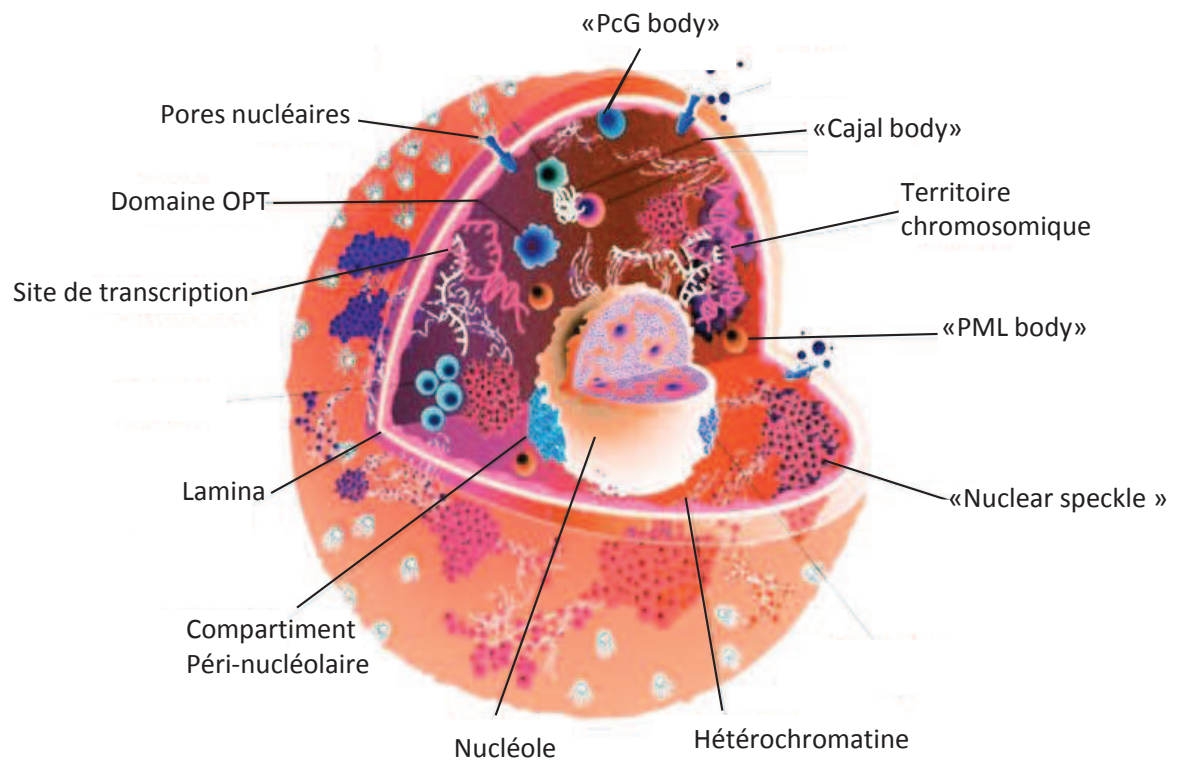


Figure 1: le noyau eucaryote.

Schéma représentant les principaux corps et domaines nucléaires.

PML:promyelocytic leukemia ; OPT: Oct1/PTF/Transcription ; PcG: polycomb group;

En réponse à des signaux environnementaux ou lors des processus permettant le développement d'un organisme multicellulaire complexe, une cellule doit pouvoir adapter son phénotype et/ou son homéostasie. Les voies de signalisation permettant l'interprétation de ces signaux aboutissent à une modification de l'expression du génome cellulaire qui passe dans la plupart des cas par une modulation plus ou moins fine de l'activité transcriptionnelle de sous-populations spécifiques de gènes (Rosenfeld *et al.*, 2006). De tels événements de régulation du transcriptome cellulaire impliquent l'intervention de nombreux mécanismes de contrôle qui doivent tenir compte de la séquence d'ADN qui est le support de l'information génétique, mais également de l'organisation spatiale et structurale de cette information au niveau du noyau des cellules eucaryotes.

I-Architecture du noyau eucaryote

D'après Theodor Schwann (1839), la cellule eucaryote est l'unité fonctionnelle de l'être vivant. A l'intérieur de chaque cellule, délimitée par une membrane, se trouve différents organites tels que la mitochondrie ou le noyau. C'est dans le noyau qu'est présente l'information génétique, portée par la molécule d'ADN. Toutes les cellules d'un même organisme, à part de rares exceptions, possèdent le même ADN dont l'expression sera adaptée selon le type cellulaire. Dans les noyaux interphasiques, il est possible de définir des domaines subchromosomaux ayant chacun des spécificités fonctionnelles (**Figure 1**).

A-L'enveloppe nucléaire

L'enveloppe nucléaire n'est pas un compartiment du noyau mais il semblerait que son existence puisse exercer un rôle dans la régulation de l'expression du génome. En effet, plusieurs études ont établi une corrélation entre un positionnement des gènes à la périphérie nucléaire, c'est-à-dire à proximité de l'enveloppe nucléaire, et leur inactivité transcriptionnelle. Une translocation des régions du génome à la périphérie nucléaire entraîne le silence des gènes présents dans cette région (Spector, 2003). Cette observation est corroborée par plusieurs autres études (Polioudaki *et al.*, 2001; de Holaska *et al.*, 2003; Somech *et al.*, 2005) qui montre que certaines protéines de l'enveloppe nucléaire comme le récepteur de la lamine B ainsi que l'emerine sont directement associées à des facteurs capables de moduler la structure chromatinienne ainsi qu'à des répresseurs transcriptionnels.

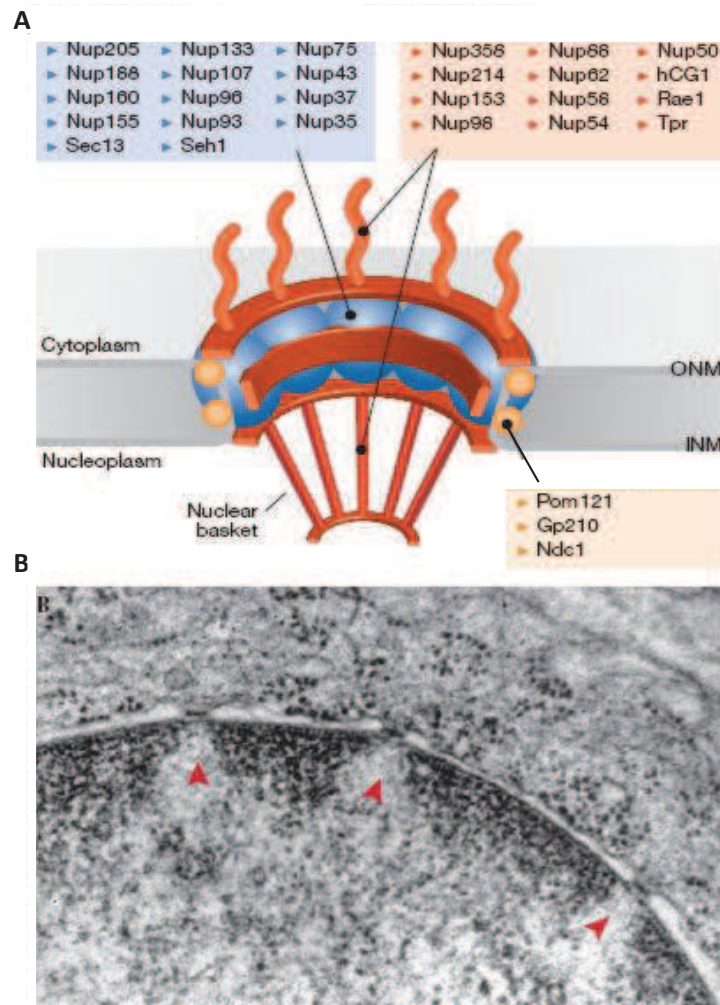


Figure 2: le pore nucléaire (Capelson et Hetzer, 2009).

(A) Illustration schématique du complexe du pore nucléaire (NPC) avec les noms des différentes nucléoporines (Nups) et protéines associées, ainsi que leurs localisations au sein du NPC.

(B) Photographie d'un noyau eucaryote observé au microscope électronique. Les flèches rouges indiquent trois pores nucléaires observables.

1-La lamina

La lamina est un constituant de l'enveloppe nucléaire présent sur la face interne, qui a un rôle essentiel au niveau de la chromatine ainsi que dans l'architecture et l'organisation des pores nucléaires (Schirmer *et al.*, 2001), dans la dissolution de l'enveloppe nucléaire au cours de la mitose puis dans son réassemblage (Lopez-Soler *et al.*, 2001), dans la réplication de l'ADN (Spann *et al.*, 1997) ainsi que sur l'expression de certains gènes dépendant de l'ARN polymérase II (ARN-PII) (Spann *et al.*, 2002) et sur la répression transcriptionnelle (Nili *et al.*, 2001). La lamina est composée de plusieurs type de lamines, les lamines B1 et B2 qui sont présentes dans les noyaux de la grande majorité des vertébrés ainsi que les lamines A et C qui apparaissent pendant ou après la différenciation cellulaire (Rober *et al.*, 1989). Les lamines ont la capacité de lier différentes protéines telles que le dimère d'histones H2A/H2B ainsi que des composants de la machinerie de transcription dépendante de l'ARN-PII (Gruenbaum *et al.*, 2003).

2-Les pores nucléaires

Tout autour de l'enveloppe nucléaire sont insérés des complexes protéiques spécialisés dans le transport bidirectionnel de macromolécules entre le noyau et le cytoplasme : les pores nucléaires (NPCs, Nuclear Pore Complexes) (**Figure 2**). Chaque pore nucléaire est constitué de plusieurs copies d'une trentaine de nucléoporines (les Nups) (Alber, 2007). Les NPCs sont des structures présentant une partie cytoplasmique, une partie transmembranaire et une partie nucléaire formant une sorte de panier. En parallèle de leur rôle dans le transport, les NPCs seraient impliqués dans l'organisation de la chromatine et la régulation génique (Akhtar et Gasser, 2007 ; Capelson et Hetzer, 2009).

Par exemple, la nucléoporine Nup2 a la capacité d'établir une barrière contre l'hétérochromatinisation d'un gène spécifique (Ishii *et al.*, 2002). De plus, des études faites sur *Saccharomyces cerevisiae* (Schmid *et al.*, 2006) ont montré qu'il existait une interaction entre les NPCs et certains gènes et cela par la méthode de ChEC ("Chromatin Endogenous cleavage"), une technique basée sur l'action d'une protéine de fusion faite d'une nucléoporine couplée à la nucléase micrococcalle, et dont les produits de digestion spécifiques sont cartographiés après hybridation sur puces à ADN. Ces interactions se feraient entre les promoteurs proximaux d'une grande partie des gènes (environ 40%) et le panier des NPCs, et cela dans une étape précoce de l'activation génique. Cette interaction pourrait être médiée par des composants de la machinerie de transcription. Deux modèles pourraient expliquer ce

recrutement de gènes actifs à la membrane, le premier appelé « bonded activation model » propose que la machinerie de transcription aurait la capacité de s'assembler grâce au panier des NPCs alors que le second, appelé « facilitated export model », serait basé sur un export plus facile des ARNms du fait d'une proximité physique entre les pores nucléaires et les ARNms en cours de transcription. Il est important de noter que ces deux modèles ne sont pas exclusifs et peuvent très bien se combiner.

Des études réalisées chez les eucaryotes supérieurs montrent au contraire l'existence d'une corrélation positive entre le positionnement des gènes au niveau de la périphérie nucléaire et leurs inactivités transcriptionnelles (Spector, 2003), même s'il semble que tous les gènes de cellules mammifères ne soient pas affectés de la même façon. En effet, des études réalisées dans le laboratoire de Wendy Bickmore (Finlan *et al.*, 2008) ont montré que la liaison d'un gène à une protéine des pores nucléaires (ici LAP2 β) réduisait fortement l'expression de ce dernier. Il semblerait que cela soit dû à une forte concentration dans cette région du noyau en histones déacétylase (HDAC) de type I et II telle que HDAC3, partenaire connu de LAP2 β (Somech *et al.*, 2005). Il semble cependant important de comprendre qu'il ne s'agirait que d'une diminution de la transcription de ces gènes et non pas d'une extinction complète. De plus, ce type de régulation ne ciblerait pas tous les gènes, certains de ceux testés conservant une expression identique quelque soit leur localisation, peut être du fait d'un contrôle par un promoteur fort. Au contraire, d'autres études ont mis en avant l'observation que Sus1, un membre du complexe d'export des ARNm (ARNm), interagit avec Nup1 et s'associe également avec le complexe histone acétyltransférase SAGA (Spt-Ada-Gcn5-Acetyltransferase) impliqué directement dans l'activation de la chromatine (Rodriguez-Navarro *et al.*, 2004). Pouvant expliquer ces rôles opposés, il apparaît que les Nups présents chez les métazoaires seraient mobiles, au contraire de leurs orthologues chez la levure. Les Nups pourraient donc se déplacer vers des régions de chromatine spécifiques et jouer un rôle au niveau de l'activation/activité transcriptionnelle de ces domaines (Capelson *et al.*, 2010), tandis que celles localisées à proximité de l'enveloppe nucléaire rendraient l'environnement plus propice à une hétérochromatinisation.

Ainsi, si plusieurs mécanismes, parfois opposés en termes de sens de modulation, semblent associer les pores nucléaires et leurs constituants à la régulation transcriptionnelle, la question de savoir si c'est la localisation en périphérie nucléaire qui est la cause d'un impact sur l'activité de la chromatine située à ce niveau ou une conséquence, reste totalement ouverte.

B-Les corps nucléaires

Il existe au sein du noyau eucaryote différents types de domaines que l'on pourrait définir comme des régions ayant une morphologie distincte de ce qui les entoure.

Dans un premier lieu, le nucléole est un sous compartiment du noyau qu'il est assez aisé d'observer *via* des colorants basiques tel que le Giemsa, mais la véritable caractérisation de son organisation a été permise par la microscopie électronique. Il est constitué d'un centre organisateur (NORs pour "Nucleolar Organizer Regions") qui correspond à des régions chromosomiques particulièrement enrichies en gènes codant pour des ARN ribosomiques (ARNr), à cela s'ajoute la présence d'ARN polymérase I, des facteurs de transcription de classe I ainsi que des topoisomérases (Strouboulis et Wolffe, 1996). Le nucléole est entouré par un composant dense fait de transcrits naissants et de la machinerie d'épissage, et par un composant granulaire incluant des ARNr matures et des sous unités ribosomales en cours d'assemblages, à différents stades (Qumsiyeh, 1999). Le compartiment péri-nucléolaire est quant à lui une structure associée mais distincte du nucléole qui contient des ARNrs transcrits par l'ARN polymérase III (Matera *et al.*, 1995). Les domaines OPT (Oct1/PTF/Transcription), d'une taille de 1 à 1.5 μ M de diamètre, sont également localisés à proximité du nucléole (Grande *et al.*, 1997) et apparaissent en phase G1 du cycle cellulaire pour disparaître ensuite en phase S. Cette région est ainsi nommée car on y détecte Oct 1, PTF (un complexe impliqué dans l'activation de la transcription de certains gènes) ainsi que de la transcription ARN-PII et PIII dépendante (Pombo *et al.*, 1998).

Quatre autres types de structure, présentant des localisations relativement indépendantes du nucléole ont également été décrites. Les "cajal bodies" ou "coiled bodies", mentionnés pour la première fois en 1903 par Cajal, sont des structures incluant de petites ribonucléoprotéines, de l'ADN et une protéine appelée la coiline. Ces structures, non délimitées par une membrane, sont impliquées dans des processus métaboliques liés aux ARNs comme la synthèse de snRNPs ("small nuclear ribonucleoproteins"), la maturation et le recyclage des ARNs ou encore la maintenance des télomères (Zhao *et al.*, 2011). Les "nuclear speckles" sont des clusters de granules mesurant de 20 à 25nm de diamètre, qui sont impliqués dans la modification et l'assemblage des facteurs d'épissage mais également de certains facteurs de transcription (Jimenez-Garcia et Spector, 1993 ; Huang et Spector, 1996). Il n'y a pas de transcription au sein de ces "speckles" mais il semblerait que l'ARN P-II y soit stockée avant son recrutement sur des sites de transcription. Enfin, les "PML bodies" n'ont pas de fonctions

précises établies même s'il est possible que cette structure ait un rôle dans la régulation transcriptionnelle, dans le stockage de protéines en vue de réguler leur concentration intranucléaire ou encore comme site de protéolyse actif (Zhong *et al.*, 2000). Dans certains types cellulaires, la présence de CBP (CREB binding protein), de p300 et de l'ARN-PII a été déterminée dans ces corps PML (promyelocytic leukemia) (La Morte *et al.*, 1998 ; Von Mikecz *et al.*, 2000 ; Boivert *et al.*, 2001).

D'un point de vue général, il semblerait que la présence de ces structures à des endroits particuliers du noyau ne soit pas due au hasard mais une conséquence directe du statut transcriptionnel de certains gènes présents à ce niveau. En effet, des études faites par Olson *et al.*, 2000 ont montré que l'introduction d'un plasmide contenant des gènes codant des ARNr entraînait la création de mini nucléoles à proximité du plasmide.

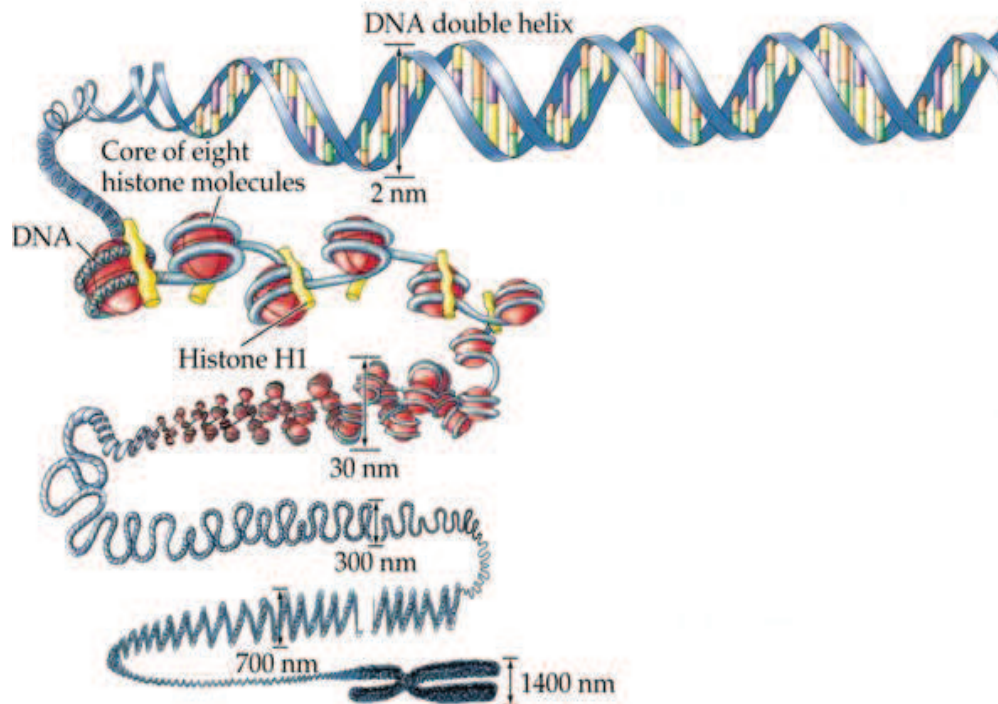


Figure 3 : Condensation de la chromatine.

La double hélice d'ADN (2nm) s'enroule autour d'un octamère d'histones et forme une fibre (11nm) de diamètre qui se replie en une structure plus condensée après inclusion de l'histone H1 (fibre de 30nm). Le degré ultime de la condensation de la chromatine est le chromosome mitotique (1400 nm).

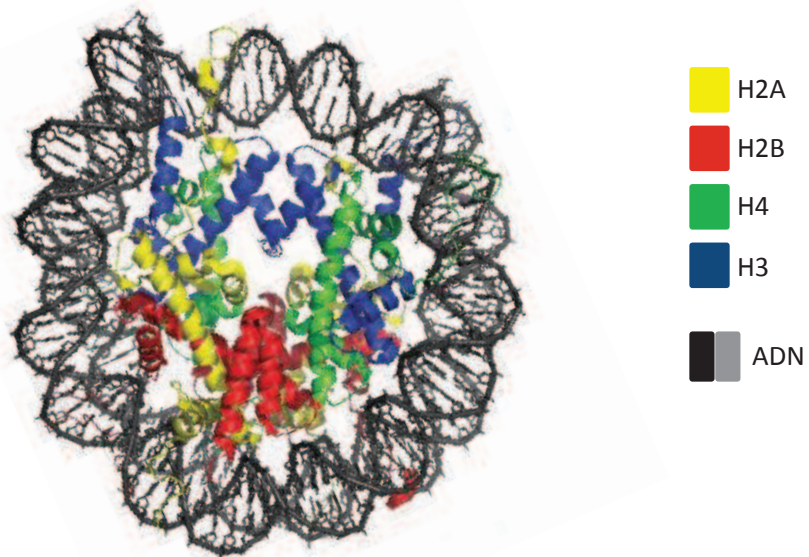


Figure 4: Structure du nucléosome (Adapté de Luger et al., 1997).

Le nucléosome est l'unité de base de la chromatine. Il est constitué de 146pb d'ADN enroulées autour d'un octamère d'histone composé de deux dimères de H2A-H2B ainsi que de deux dimères de H3-H4.

II-La chromatine

A-Organisation structurale de la chromatine

1-Le collier de perles

Au sein du noyau, l'ADN n'est pas nu mais organisé sous forme de chromatine : 146 paires de bases (pb) de la double hélice de 2nm sont enroulées autour d'un octamère d'histones (H2A, H2B, H3 et H4 présentes chacune en double exemplaires). Cette structure appelée le nucléosome est l'unité de base répétée de la chromatine (fibre de 11nm) (Wolffe, 1998) et la distance relative entre chaque nucléosome influe sur le degré de condensation de la chromatine et donc sur l'accessibilité de l'information génétique aux facteurs de transcription (Felsenfeld, 1992). L'inclusion de l'histone H1 au sein de cette structure permet une condensation plus importante de la chromatine (fibre de 30nm) (Misteli *et al.*, 2000), dont le degré ultime de compaction est représenté au sein des chromosomes mitotiques (14000nm) (**Figure 3**).

2-Les histones

Les histones sont de petites protéines basiques très conservées au cours de l'évolution dont la structure peut être divisée en trois domaines (Luger et Hansen, 2005). Le domaine central des histones est une région très conservée lors de l'évolution, comme en atteste l'histone H4 qui présente une différence de seulement 2 acides aminés (A.A) entre le thymus de veau et le petit pois (DeLange *et al.*, 1969). Ce domaine est également très structuré, présentant une organisation en trois hélices alpha séparées de deux boucles (Arents et Moudrianakis, 1995). Les deux autres domaines des histones correspondent à leurs extrémités carboxy et amino terminales (appelé par la suite C-terminal et N-terminal, respectivement). Au contraire du domaine central, ces "queues" des histones ne sont pas structurées et varient dans leur composition ; elles sont le lieu privilégié des modifications post traductionnelles des histones dont la complexité et l'importance sera développée dans un chapitre dédié par la suite. Dans le noyau, les histones H3 et H4 s'associent en hétérodimères tout comme les histones H2A et H2B, le tout formant à terme un octamère (Tagami *et al.*, 2004) (**Figure 4**) et non pas en un tétramère (H3-H4) avec deux dimères H2A-H2B comme définit il y a quelques années (Wolffe, 1998).

3-Les histones de liaison

L'histone de liaison H1 (et son variant H5) est présent tout le long de la chromatine du génome eucaryote. Il est nécessaire pour le repliement de la chromatine en une structure plus élaborée, plus compacte: la fibre de 30nm (Huynh *et al.*, 2005) qui est moins accessible aux enzymes de remodelage de la chromatine et même réfractaire à la transcription (Horn *et al.*, 2002). Il existe 11 sous types de H1 (Clausell *et al.*, 2009); 7 étant somatiques, 3 retrouvés dans le sperme et 1 spécifique des oocytes. Les histones de liaison sont de petites protéines basiques, composées comme les précédentes de trois domaines: un domaine globulaire assez bien conservé entre les espèces, une extrémité N-terminale de 20 à 35 A.A et une extrémité C-terminale d'environ 100 A.A, stabilisant le repliement de la chromatine (Caterino et Hayes, 2011). Malgré la présence d'H1 sur l'ensemble du génome, quelque soit le type de chromatine (hétérochromatine ou euchromatine), certaines régions dépourvues de cet histone ont été identifiées, principalement au niveau de régions denses en gènes actifs et particulièrement au niveau du site de début de la transcription (TSS) de ces gènes. Cependant, 10% des sites ne présentant pas H1 semblent être localisés au niveau de régions intergéniques ; ces sites ou une partie d'entre eux pourraient constituer des éléments régulateurs (Braunschweig *et al.*, 2009). Il semblerait donc que H1 fasse partie des modulateurs majeurs de la chromatine.

4-Les variants d'histones

Des variants d'histones ont été découverts chez plusieurs espèces eucaryotes, différant de façon plus ou moins importantes des protéines histones conventionnelles au niveau de leurs séquences (Malik et Henikoff, 2003) mais également dans leur assemblage au sein des nucléosomes. Ces différences ont une répercussion au niveau de la fonction de ces variants, et présentent des rôles spécifiques dans de nombreux processus biologiques, tels que l'organisation centromérique, la réparation de l'ADN, l'inactivation du chromosome X (Xi) ou encore la transcription.

a-Les variants de H2A

H2A.Z est un variant de l'histone H2A dont l'inactivation chez la Souris est létale (Faast *et al.*, 2001). Les nucléosomes contenant ce variant présentent des caractéristiques biochimiques et physiques spécifiques et sont réfractaires à certaines modifications post-traductionnelles. H2A.Z possède des rôles variés ; il est notamment impliqué dans la régulation de la transcription génique, pouvant servir de frontière à la diffusion de

l'hétérochromatine, et intervient dans le développement embryonnaire (Guillemette et Gaudreau, 2006 ; Eirin-Lopez et Ausio, 2007; Kusch et Workman, 2007). Dans l'euchromatine, H2A.Z serait présent uniquement au niveau de la région promotrice de la sous-population des gènes transcrit par l'ARN-Pol II ne présentant pas de TATA box. Ce variant serait "déposé" à ce niveau par l'intermédiaire du complexe SWR1 (Swi/Snf related protein) chez la levure ou son homologue (swr1) chez l'humain (Zhang *et al.*, 2005). Un recrutement alternatif d'H2A.Z est envisageable dans la mesure où des études effectuées au sein du laboratoire de Madhani (Raisner *et al.*, 2005) ont indiqué que H2A.Z serait également présent sur le promoteur de gènes actifs ou inactifs présentant un motif d'ADN spécifique consistant en deux segments redondants de 10 pb contenant le site de liaison de la protéine Reb1 suivi d'une séquence de 7 pb : dT:dA. Les gènes cibles de H2A.Z présentent une région libre en nucléosomes, située environ 200 pb en amont du site de transcription (TSS), cette région étant entourée par deux nucléosomes H2A.Z (Raisner *et al.*, 2005; Yuan *et al.*, 2005). D'autre part, H2A.Z est préférentiellement lié aux promoteurs de gènes réprimés lors de changements importants du transcriptome chez la levure à la suite d'un "heat shock", une perte de H2A.Z étant observée sur les promoteurs de gènes actifs associée à un gain de ce variant au niveau des promoteurs inhibés (Zhang, 2005). Il est possible que cette perte en H2A.Z soit associée à une moins grande stabilité des nucléosomes contenant ce variant par rapport à ceux contenant H2A. Au niveau de l'hétérochromatine, H2A.Z aurait deux rôles. Tout d'abord, sa présence à proximité des télomères permettrait le maintien de frontières entre l'hétérochromatine des télomères et les gènes adjacents, en relation avec l'acétylation de l'histone H4. En effet, chez la levure, une absence d'H2A.Z ou bien un défaut d'acétylation d'H4 entraîne une diffusion des protéines Sir, impliquées dans l'hétérochromatisation, des télomères vers les gènes adjacents (Zhang *et al.*, 2004). Ce rôle a été confirmé chez le Poulet (Bruce *et al.*, 2005). Deuxièmement, H2A.Z aurait un rôle fondamental au cours du développement précoce de l'embryon. Ceci résultant d'un fort enrichissement de ce variant au sein de l'hétérochromatine constitutive où il interagit avec HP1 α et INCENP, deux protéines importantes pour la ségrégation chromosomique (Rangasamy *et al.*, 2003). Il a également été montré dans les cellules ES que H2A.Z est enrichi sur de nombreux gènes en lien avec le développement comme en atteste l'annotation en termes GO (Gene Ontology) des gènes cibles d'H2A.Z qui correspondent majoritairement à des facteurs de transcription ayant des rôles importants dans les processus de développement comme les facteurs FOX ou GATA (Creyghton *et al.*, 2008).

Le variant de type **MacroH2A** existe sous 3 formes, MacroH2A1.1, MacroH2A1.2 et MacroH2A.2. La partie N-terminale de MacroH2A.1 est à 64% identique à celle de H2A mais sa partie C-terminale est très différente. De très grande taille, elle constitue une région non histone (Pehrson et Fried, 1992) qui contient un domaine appelé Macro ayant la capacité à lier des ADP-ribose (Adénosine diphosphate-ribose) ainsi que d'autres petites molécules associées (Kustatscher *et al.*, 2005), mais sa fonction est pour le moment encore inconnue. MacroH2A est particulièrement enrichi sur le Xi dans les cellules mammifères (Ladurner, 2003) ainsi qu'au niveau de certains gènes réprimés en attente d'activation (Buschbeck *et al.*, 2009). Une étude récente, dirigée par le laboratoire de Lee Kraus (Gamble *et al.*, 2010) a cependant montré une localisation de macroH2A en amont du TSS de certains gènes impliqués dans la régulation du développement et des cascades de signalisation. De plus, dans certains cas, MacroH2A aurait une capacité à réguler positivement la transcription, ce qui étendrait donc les rôles historiques de ce variant généralement associé à des régulations négatives.

H2AX est un variant de l'histone H2A découvert dans les années 1980 qui possède une structure similaire à H2A excepté un motif présent sur son extrémité C-terminale (Redon *et al.*, 2002). H2AX est impliqué dans la réparation des cassures doubles brins de l'ADN, la phosphorylation de sa sérine 139 permettant en effet le recrutement de nombreuses protéines impliquées dans la réponse à ce type de dommages de l'ADN (Paull *et al.*, 2000).

b-Les variants de H3

H3.3 est un variant de l'histone H3 (Hake et Allis, 2006) spécifiquement enrichi sur les gènes ayant un fort taux en CpG au niveau des promoteurs (HCP), qu'ils soient transcriptionnellement actifs ou non, au niveau de leurs séquences régulatrices, de leurs corps et après leurs signaux de terminaison; et cela dans les cellules souches embryonnaire (cellules ES) ainsi que dans les cellules différenciées. Cependant, il est important de noter que H3.3 est perdu par les gènes pluripotents tel que Nanog ou Oct 4 au cours de la différenciation cellulaire. H3.3 est incorporé dans la chromatine par deux voies distinctes. La première, nécessitant la protéine chaperonne HIRA, est mise en jeu sur les gènes HCP alors que la seconde, impliquant Atrx est relative au chargement de H3.3 au niveau des télomères (Glodberg *et al.*, 2010). Il a également été établie une anti-corrélation entre la présence du variant d'histone H3.3 (localisé majoritairement au niveau de régions transcriptionnellement actives) et celle de H1 au niveau de la chromatine, probablement afin que les gènes et les



Figure 5: Exemples de modifications post-traductionnelles des histones.

Les extrémités N-terminales des histones sont les cibles de modifications post-traductionnelles, telles que l'acétylation et la méthylation des résidus lysines et arginines, la phosphorylation sur les sérines et les thréonines ou encore l'ubiquitination.

éléments régulateurs restent accessibles à divers facteurs impliqués dans la régulation transcriptionnelle (Braunschweig *et al.*, 2009).

La protéine **CENP-A** est un variant de l'histone H3 chez les vertébrés. On la trouve uniquement au niveau des centromères où elle permet la formation d'un nucléosome spécifique dont la structure reste à déterminer (Henikoff et Ahmad, 2000). Deux études ont montré le remplacement de l'histone H3 par CENP-A au niveau des centromères, dont des études de ChIP-chip (immunoprécipitation de la chromatine et analyse sur puce) qui montrent une déplétion en H3 *in vivo* au niveau de domaines où CENP-A est présente (Lo *et al.*, 2001), mais également des études de microscopie chez la *Drosophila* mettant en avant une absence de H3-GFP (Green fluorescent protein) au niveau des domaines de liaison de CID (orthologue de CENP-A) (Ahmad et Henikoff, 2001). Paradoxalement, d'autres travaux montrent au contraire qu'il existerait une alternance de nucléosomes incluant CENP-A avec d'autres présentant H3 au sein du centromère (Blower *et al.*, 2002). CENP-A est essentielle dans l'organisation et la fonction du kinétochore, sa déplétion par ARN interference (ARNi) chez *C.elegans* (Buchwitz *et al.*, 1999) ou chez *D.melanogaster* (Blower et Karpen, 2001) ou bien son inactivation par KO chez la Souris (Howman *et al.*, 2000) entraînant d'importants problèmes dans la ségrégation des chromosomes et même un phénotype létal dans certains cas.

B-Modifications de la chromatine

1-Les modifications des histones

Les histones subissent de très nombreuses modifications post-traductionnelles telles que l'acétylation et la méthylation de leurs résidus lysines ou arginines, la phosphorylation sur les sérines et les thréonines ainsi que l'ubiquitination, la sumoylation, l'ADP-ribosylation ... (Peterson et Laniel, 2004) (**Figure 5**). De plus, un niveau de complexité supérieur de ces modifications est atteint par la possibilité d'accepter pour chaque résidu lysine, un, deux ou trois groupement(s) méthyle et pour chaque arginines un ou deux groupement(s) méthyle. Ces modifications sont mises en place par l'intermédiaire de protéines dont la première à avoir été découverte est une histone acétyl-transférase (HAT), chez la levure en 1996 (peterson et Laniel, 2004). Depuis, un certain nombre de coactivateurs transcriptionnels comme CBP/p300 ont été identifiés comme ayant une activité HAT (Couture et Trievel, 2006). De la même

façon, plusieurs co-répresseurs de l'activité transcriptionnelle ont été mis en avant comme Rpd3 et présentent une activité HDAC. Les enzymes capables de modifier les histones sont actuellement classées par famille avec entre autres les HAT, les HDAC, les histones méthyltransferase (HMT) et les kinases.

Au niveau du génome, l'apposition de marques post-traductionnelles sur les histones se fait en des sites précis, ce ciblage étant principalement dû à l'effet combiné des enzymes de modification des histones et de régulateurs de certaines fonctions biologiques comme des régulateurs transcriptionnels. Nous pouvons prendre l'exemple du complexe SAGA qui possède une sous unité présentant une activité HAT, ayant la capacité à interagir avec des protéines impliquées dans l'activation de gènes spécifiques. Ce sont ces interactions qui ciblent SAGA sur une région donnée. Cependant, pour certains phénomènes comme la réparation de l'ADN, le ciblage de modifications post-traductionnelles d'histones serait plus direct, la lésion de l'ADN provoquant le recrutement direct de la kinase ATM (ataxia telangiectasia mutated) sur le site de cassure double brin dans le but de phosphoryler l'histone H2A ou son variant H2AX (Peterson et Laniel, 2004).

Les modifications post-traductionnelles des histones peuvent avoir un impact direct sur l'organisation locale de la chromatine en altérant les interactions histones/ADN. Par exemple, la queue de l'histone H3 présentant 13 A.A avec une charge positive, l'acétylation d'1/4 de ces résidus suffirait à diminuer de 10 à 30% les charges positives, et donc à perturber ses interactions avec des molécules chargées négativement comme l'ADN ou d'autres protéines. (Peterson et Laniel, 2004). Les modifications post-traductionnelles des histones peuvent également avoir un impact transcriptionnel indirect en modulant le recrutement de certaines protéines capables de les reconnaître (Martin et Zhang, 2005 ; Couture et Trievel, 2006). Nous pouvons prendre divers exemples comme les protéines à bromodomaine, qui se lient spécifiquement aux lysines acétylées : l'acétylation de la lysine 9 de l'histone H4 (H4K9ac) permet de recruter le complexe de remodelage de la chromatine ATP-dépendant, SWI/SNF (switch/sucrose non fermentable) *via* le bromodomaine de sa sous unité Brg1. En ce qui concerne les résidus méthylés, ceux-ci sont reconnus par des protéines présentant des domaines spécifiques : les chromodomains, les domaines Tudor qui peuvent se lier aux lysines mais également aux arginines méthylées et enfin les domaines à répétition WD40 qui forment une hélice β et qui sont inclus dans des protéines impliquées dans de multiples processus biologiques comme la régulation transcriptionnelle, l'épissage des ARNs ou encore

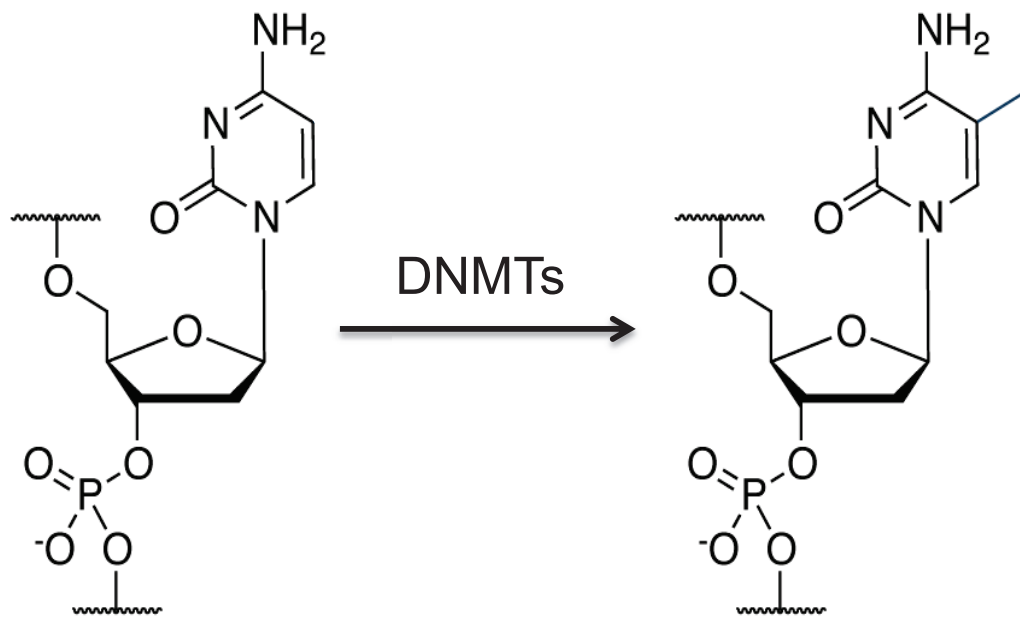


Figure 6: Conversion de la cytosine en 5-méthylCytosine (5mC).

Au sein du génome, peu de modifications affectent directement l'ADN. La mieux caractérisée est la méthylation des cytosines en méthyl-cytosine ou 5mC (appelée cinquième base de l'ADN) par l'intermédiaire des DNMTs.

le trafic cellulaire. Les histones méthylés peuvent également selon leur localisation et leur nombre de résidus méthyles, recruter d'autres protéines comme HP1 (hétérochromatin protein 1) ou les protéines du groupe Polycomb, impliquées dans la compaction de la chromatine (Margueron *et al.*, 2005)

De par ces rôles directs ou indirects dans la modulation de la structure/organisation chromatinienne, les modifications post-traductionnelles des histones sont impliquées dans tous les processus impliquant l'ADN : transcription des gènes, inactivation du X, formation de l'hétérochromatine, division cellulaire, réplication et réparation de l'ADN (Peterson et Laniel, 2004 ; Ho et Crabtree, 2010). De nouvelles modifications sont découvertes très régulièrement et il est possible que chaque résidu accessible soit ciblé par une modification, définie ou non pour le moment. L'ensemble de ces « marques d'histones » peut être à l'origine d'un code histone (Strahl et Allis, 2000), au sein duquel certaines modifications au niveau de sites spécifiques pourraient être reliées à des fonctions biologiques particulières. Par exemple, l'acétylation de la lysine 9 de l'histone H3 (H3K9ac) ou la triméthylation de la lysine 4 de l'histone H3 (H3K4me3) sont reliées à l'activation transcriptionnelle alors que la triméthylation de la lysine 9 de l'histone H3 ainsi que l'hypoacétylation des histones H3 et H4 sont plus reliées à la répression transcriptionnelle (Li *et al.*, 2010).

2-Les modifications de l'ADN

1-Méthylation des cytosines.

Présentation

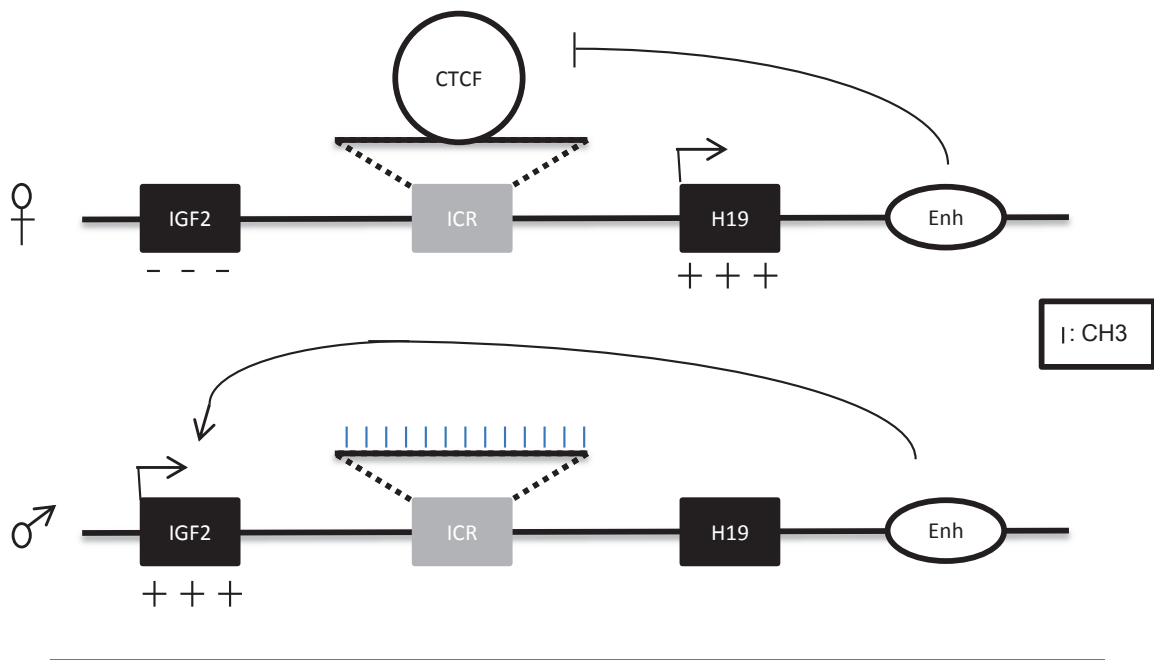
Parmi le peu de modifications connues touchant à l'ADN, la mieux caractérisée est certainement sa méthylation au niveau du carbone 5 des cytosines (5m-C ; surnommée la cinquième base de l'ADN) (**Figure 6**) et cela dans le contexte d'un dinucléotide CG : 5m-CpG. Cette modification est considérée comme étant épigénétique, dans le sens où elle peut être héritable (transmission à travers les générations de cellules). Les CpG sont sous-représentés dans le génome humain et ne sont pas répartis uniformément dans le génome. Ils sont principalement enrichis dans de petites séquences riches en CpG : les îlots CpG que l'on trouve à 60% dans les promoteurs des gènes chez la Souris ou l'Homme. Ces îlots sont également enrichis dans le premier exon des gènes ainsi que dans des régions à proximité de l'extrémité 3' des gènes (Jones et Takai, 2001). La méthylation n'est pas une modification existant chez tous les eucaryotes, elle est en effet absente du génome de *C.elegans* (Bird,

2002) et très faible chez la Drosophile, où elle est ciblée sur le dinucléotide CpT (Gowher *et al.*, 2000 ; Lyko *et al.*, 2000) et non pas CpG comme chez les vertébrés. La 5m-C représente dans le génome des mammifères seulement 1% des bases du génome mais au niveau des CpG, on observe une augmentation du niveau des bases méthylées avec 70 à 80% des CpG qui sont méthylées. Les études de la méthylation de l'ADN à grande échelle considérant l'ensemble du génome ont été facilitées par la mise en place de la technique du bisulfite-seq (Clark *et al.*, 1994), dont le principe fait intervenir une conversion des cytosines non méthylées en uracile suivi d'une étape de séquençage haut débit. Cette technique a permis de mettre en avant l'existence, au sein des cellules non différenciées, d'évènements de méthylation très minoritaires sur d'autres types de dinucléotides: les CHG et les CHH où H représente soit un A, soit un C soit un T. Cependant, cette méthylation en dehors des CpG est perdue au cours de la différenciation cellulaire (Lister *et al.*, 2009). La technique du bisulfite est difficile à mettre en œuvre et il est techniquement compliqué d'identifier les cibles de la méthylation par séquençage. Afin de simplifier ces études, le laboratoire de D. Schubeler a mis au point une autre technique, le MeDIP (immunoprécipitation des cytosines méthylées) qui permet *via* un anticorps dirigé contre la 5m-C, d'immunoprécipiter des séquences méthylées. Le MeDIP peut être analysé sur puce ou bien par séquençage (Weber *et al.*, 2005).

Mise en place et maintien de la méthylation

Lors du développement embryonnaire, une phase de déméthylation globale du génome a été observée. Elle est suivie par une phase de méthylation *de novo* et cela dans le but de déterminer le patron de méthylation des cellules somatiques. Le résidu méthyle est apposé par des enzymes appelées DNMT (« DNA methyl transferase ») qui sont au nombre de 3 chez les mammifères à avoir cette capacité : DNMT1, DNMT3a et DNMT3b. La maintenance de la méthylation au cours de la réplication n'est pas simple. Elle est principalement permise par l'aspect semi conservatif de la réplication de l'ADN où DNMT1 méthyle le nouveau brin en fonction du brin parental (Pradhan *et al.*, 1999). Il semblerait cependant qu'il existe un autre mécanisme à l'origine du maintien de la méthylation au cours de la réplication puisqu'un maintien de la méthylation dans un contexte pauvre en DNMT1 a été observé (Jaenisch, 1997). DNMT3a et 3b sont quant à elles impliquées principalement dans des évènements de méthylation *de novo* (Okano *et al.*, 1998a). Elles sont très exprimées dans les cellules embryonnaires, au moment où on observe la phase de re-méthylation du génome. Deux modèles se font face en ce qui concerne le choix des séquences à méthyler *de*

A



B

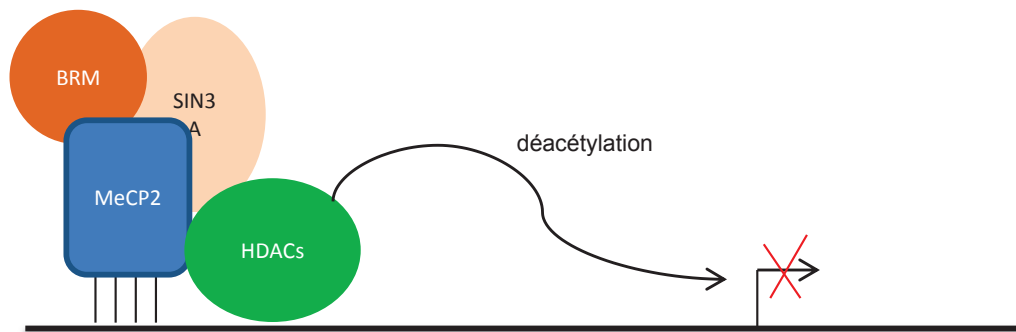


Figure 7: Exemples de rôles de la méthylation.

(A) La méthylation peut empêcher, de façon directe, la liaison d'un facteur donné. C'est le cas sur le locus *H19/IGF2*. Sur l'allèle maternel, une séquence régulatrice de ce locus, l'ICR (« Imprinting Control Region ») est liée par CTCF ce qui empêche un enhancer situé en aval des deux gènes d'activer *IGF2*. Au contraire, au niveau de l'allèle paternel, l'ICR est méthylé, empêchant ainsi la liaison de CTCF. L'enhancer a donc la capacité d'activer *IGF2*.

(B) La méthylation peut également permettre le recrutement de protéines de la famille MBD comme MeCP2. Cette dernière va à son tour recruter des cofacteurs comme SIN3A ou des HDACs, responsables d'une hétérochromatinisation de la chromatine environnante et donc d'une répression des gènes méthylés.

nov. Le premier est un modèle par défaut dans lequel la méthylation serait un processus affectant potentiellement tous les CpG. Mais ce modèle semble inapproprié. En effet, le fait que la DNMT3b est ciblée sur des séquences précises comme des séquences d'ADN répétitif et des îlots CpG présents sur le Xi (Miniou *et al.*, 1994) est en faveur du second modèle où la méthylation serait ciblée sur des CpGs particuliers.

Rôles de la méthylation

L'implication de la méthylation des cytosines a été démontrée dans diverses fonctions dans l'organisme adulte telles que la régulation de la transcription, la sauvegarde de l'intégrité du génome, la structuration et l'organisation du génome ainsi que dans la réplication. La présence de méthylation sur les séquences répétitives de types satellites, SINEs (short interspersed nuclear elements), LINEs (long interspersed nuclear elements) permettrait de les réprimer (Zemach et Zilberman, 2010). On peut prendre l'exemple des SINEs dont les plus importants au sein du génome sont la famille des séquences Alu. Une déméthylation artificielle de ces séquences est suffisante pour promouvoir leurs fonctions régulatrices (Wodcock *et al.*, 1997). La méthylation de l'ADN peut permettre une répression transcriptionnelle par deux mécanismes distincts. Le premier fait intervenir une interférence directe du groupement méthyl pour la liaison d'un facteur de transcription sur sa séquence cible. C'est par exemple le cas au niveau du locus H19/IGF2 soumis à l'empreinte parentale, où la liaison de la protéine CTCF sur l'allèle paternel est inhibée du fait de la présence de méthylation, permettant ainsi l'expression du gène *Igf2* (Bell et Felsenfeld, 2000) (**Figure 7A**). Le second mode d'action implique au contraire des protéines ayant une affinité spécifique pour le 5m-C, comme les protéines de la famille MBD (Methyl Binding Domain). Les 5 protéines appartenant à cette famille présentent un domaine de liaison aux 5m-C. Quatre d'entre elles (MBD1, MBD2, MBD3 et MeCP2) sont impliquées dans la répression de la transcription *via* la méthylation (Bird et Wolffe, 1999). En ce qui concerne MeCP2 par exemple, elle contient, en plus de son MBD, un domaine de répression transcriptionnelle: le TRD (Transcriptional Repression Domain) qui a la capacité à interagir avec la protéine SIN3A, elle-même en interaction avec des HDAC dont le rôle de déacétylation des histones permet l'établissement d'un environnement chromatinien répressif (Knoepfler et Eisenman, 1999) (**Figure 7B**). Un tel lien entre les modifications de l'ADN et celles des histones a également été montré, MBD2 étant présent au sein d'un complexe multiprotéique comprenant les HDAC1 et HDAC2 (Ng *et al.*, 1999).

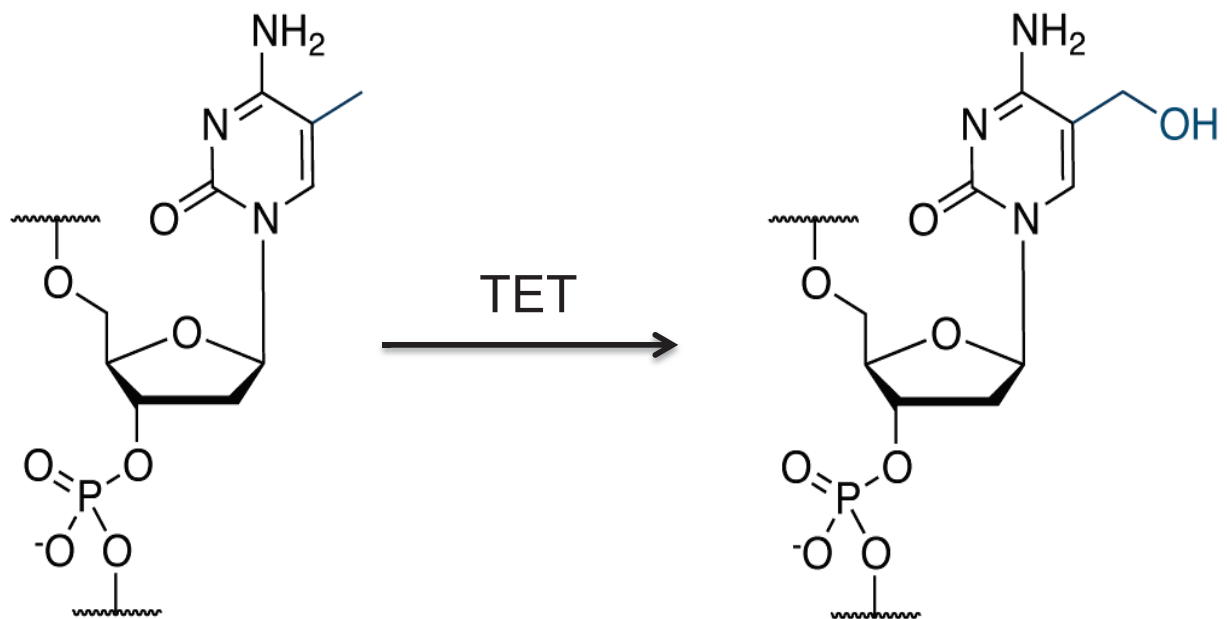


Figure 8: Conversion de la 5-méthylcytosine (5mC) en 5-hydroxyméthylcytosine (5hmC).

La 5hmC, considérée comme la sixième base de l'ADN, serait issue d'une hydroxylation de la 5mC par des protéines de la famille TET

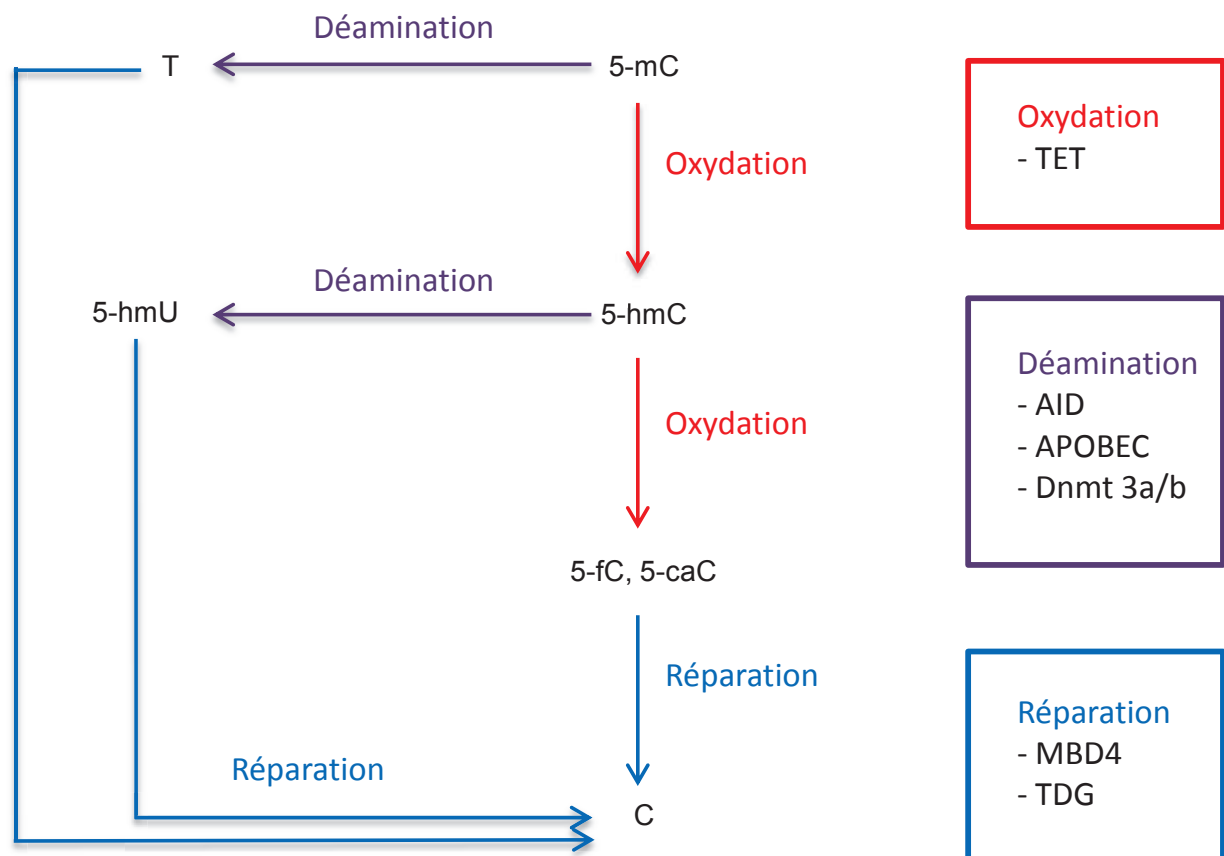


Figure 9: Mécanismes de déméthylation active des cytosines.

La déméthylation active des cytosines peut se faire via deux voies, la première ayant pour origine la déamination de la 5-mC en T par des protéines de type AID/APOBEC ou même les Dnmt3. Des études récentes ayant montré l'incapacité des protéines de la famille AID/APOBEC à prendre en charge la 5-mC, cette voie de déamination pourrait également nécessiter une étape initiale d'oxydation de la 5-mC en 5-hmC puis une déamination de cette dernière en 5-hmU. Dans un cas comme dans l'autre le mismatch T/G ou 5-hmU/G serait ensuite reconnu par le système de réparation de base de l'ADN (BER) menant au remplacement de la base incorrecte en cytosine. La seconde voie repose sur une cascade d'oxydations par les protéines TET, partant de la 5-mC et aboutissant à 5-fC puis à la 5-caC en passant par la 5-hmC. La dernière étape est une reconnaissance des bases 5-caC par le système BER qui les corrige en C.

5-mC: 5-méthylcytosine, 5-hmC: 5-hydroxyméthylcytosine, T: Thymine, 5-hmU: 5-hydroxyméthyluracile, 5-caC: 5-carboxylcytosine, 5-fC: 5-formylcytosine, TET: Ten Eleven Translocation, AID: Activation Induced Deaminase, APOBEC: apolipoprotein B mRNA editing enzyme catalytic subunit, DNMT3a/b: DNA MethylTransferases 3a/b, MBD4: Methyl-CpG Binding Domain protein 4, TDG: Thymine DNA Glycosylase.

b-Hydroxyméthylation des cytosines

Une seconde modification de l'ADN a été (re)découverte récemment au même moment par deux laboratoires différents : la 5-hydroxyméthylcytosine (5-hmC) (Kriaucionais et Heintz, 2009 ; Tahiliani *et al.*, 2009). Cette modification, considérée comme la sixième base de l'ADN, serait issue d'une hydroxylation de la 5-mC par les protéines de la famille TET (ten-eleven translocated) (Tahiliani *et al.*, 2009) (**Figure 8**). La 5-hmC est assez rare (4% des cytosines dans un contexte CpG) et ne serait présente que sur les bras des chromosomes mais pas au niveau de la chromatine péri-centromérique et de ses régions répétées (Szulwach *et al.*, 2011).

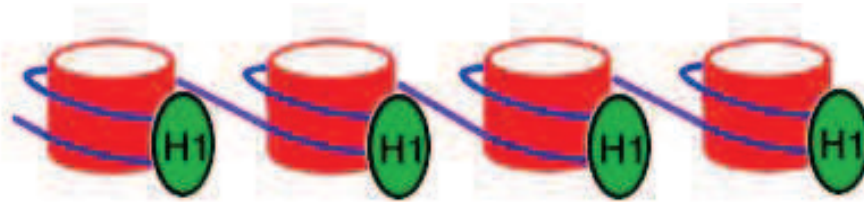
La localisation de la 5-hmC est très spécifique avec un enrichissement au niveau du corps des gènes et principalement dans les exons et autour du TSS. La distribution de la 5-hmC au niveau des promoteurs proximaux des gènes est dépendante de leurs niveaux d'expression. En effet, les gènes moyennement exprimés présentent une distribution bi-modale autour du TSS ainsi qu'un enrichissement dans le corps des gènes, contrairement aux gènes peu exprimés qui eux sont enrichis en 5-hmC au niveau de leurs TSSs mais pas dans leurs corps. Cette distribution peut être mise en relation avec certaines marques d'histones telles que H3K4me1 et K4me2 qui montrent un enrichissement identique à celui de la 5-hmC dans les gènes moyennement exprimés. En ce qui concerne les enhancers, la présence de certaines modifications post-traductionnelles d'histone comme H3K27ac ou encore H4K5ac sont des pré requis pour la présence de la 5-hmC (Szulwach *et al.*, 2011).

Il est possible que l'hydroxyméthylation de l'ADN influe sur la structure de la chromatine et de cette façon sur l'activité transcriptionnelle locale en permettant le recrutement de protéines reconnaissant spécifiquement les 5-hmC ou bien par exclusion des protéines reconnaissant les 5-mC telle que MeCP2, incapable de se lier à une séquence hydroxyméthylée (Tahiliani *et al.*, 2009). Cependant, la 5-hmC n'est pas strictement corrélée à une activité transcriptionnelle (Wu *et al.*, 2011).

c-Dynamiques des modifications de l'ADN

La méthylation n'est pas une caractéristique statique du génome, elle est au contraire particulièrement dynamique et nécessite donc des mécanismes de déméthylation (**Figure 9**). Il existe plusieurs exemples de déméthylation active, comme dans le génome paternel après fertilisation (Mayer *et al.*, 2000), ou bien lors de la formation des cellules primordiales de la

Hétérochromatine



Euchromatine

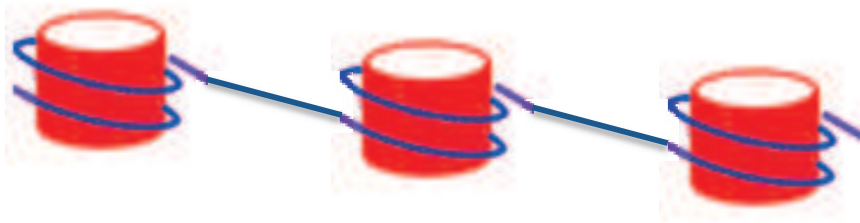


Figure 10: Etats de la chromatine.

La chromatine peut exister sous deux formes au sein des noyaux: l'hétérochromatine qui correspond à une compaction forte de la chromatine par l'insertion de l'histone de liaison H1, et l'euchromatine qui présente une structure moins condensée et plus riche en gènes.

En rouge les histones, en bleu l'ADN et en vert l'histone de liaison H1.

lignée germinale dans les embryons (Hajkova *et al.*, 2002). Le premier mécanisme à avoir été proposé est une déamination de la 5-mC par les protéines de la famille AID (Activation induced deaminase)/APOBEC (Apolipoprotein B mRNA editing enzyme) (Conticello, 2008) en thymine. Cette substitution serait corrigée par l'intervention successive de glycosylases spécifiques des mésappariements T:G, telles que la TDG (Thymine DNA glycosylase) ou MBD4 chez les mammifères (Millar *et al.*, 2002 ; Hardeland *et al.*, 2003), puis par la reconnaissance du site abasique ainsi créé par le système de réparation BER qui permet l'incorporation d'une cytosine non méthylée (Fritz et Papavasiliou, 2010). Cependant, des études plus récentes semblent mettre en avant une incapacité des protéines AID/APOBEC à déméthyliser la 5-mC (Guo *et al.*, 2011). Ce travail montre la nécessité d'une étape initiale d'hydroxylation de la 5-mC en 5-hmC. La sixième base de l'ADN serait alors reconnue par les déaminases de la famille AID/APOBEC (principalement APOBEC1 et APOBEC2) (Morgan *et al.*, 2004 ; Rai *et al.*, 2008) donnant ainsi de la 5-hydroxyuracile, prise en charge par des protéines du système de réparation BER (base excision repair) (Gua *et al.*, 2011). Un second mécanisme mettant en œuvre une cascade d'oxydation par les protéines TET serait également possible, la 5-mC étant hydroxylée successivement en 5-hmC, 5-formylcytosine (5-fC) puis en 5-carboxylcytosine (5-caC), ces 2 dernières étant reconnues par le système de réparation BER (Ito *et al.*, 2011).

C-Organisation fonctionnelle et condensation de la chromatine

Il est communément admis depuis un certain nombre d'années que l'euchromatine, qui correspond à de la chromatine en partie décondensée, contiendrait des gènes particulièrement actifs, contrairement à l'hétérochromatine qui serait de la chromatine transcriptionnellement inactive. Il apparaît cependant que cette idée est une simplification de la réalité.

1-L'hétérochromatine

Au début des années 60, Mirsky et ses collègues découvrent en périphérie nucléaire une chromatine très condensée : l'hétérochromatine (**Figure 10**). Cette compaction peut être mise en relation avec une structure globalement répressive à tout événement de régulation. On peut cependant sous-diviser ce type de chromatine en hétérochromatine constitutive qui est transcriptionnellement silencieuse en permanence dans une cellule différenciée et en hétérochromatine facultative qui contient des gènes temporairement réprimés.

L'hétérochromatine représente la majeure partie du génome mammifère (environ 96%) et inclut majoritairement des séquences non codantes ou répétitives.

a-Mécanismes généraux d'hétérochromatinisation

L'établissement d'une structure hétérochromatinienne est associée à l'intervention de l'histone de liaison H1 et/ou à des modifications de l'état de la chromatine faisant intervenir une protéine spécifique, HP1 ou/et des protéines du groupe polycomb. Parmi les autres mécanismes permettant l'hétérochromatinisation, il a été récemment mis en évidence un lien avec ce que l'on appelle la transcription non spécifique. En effet, la présence d'ARNs double brin, caractéristique de ce phénomène, est reconnue par le complexe DICER qui les transforme en siRNAs (small interference RNA), permettant ainsi le ciblage de la région incriminée par le complexe protéiques RITS, aboutissant à la déacétylation d'histones, la méthylation de certaines lysines, le recrutement d'HPA et enfin à l'établissement de la méthylation de l'ADN. Le KO des facteurs impliqués dans le processing des siRNAs entraîne des défauts sur la triméthylation d'H3K9 mais également sur la liaison de HP1. Chez la Levure, l'hétérochromatinisation ne ferait pas intervenir ce mécanisme, mais débiterait au niveau de sites spécifiques du génome, se diffusant ensuite le long de la chromatine en inhibant toute la région (Renauld *et al.*, 1993). Ce mécanisme implique l'intervention de protéines Sir et plus exactement Sir2, Sir3 et Sir4, qui, en interagissant physiquement avec la chromatine, la rendent ainsi inaccessible à la liaison d'autres facteurs tels que la RNA-Pol II. En général, ce type de diffusion de l'hétérochromatine est arrêté par la présence de frontières de chromatine, assez bien définies chez la levure. Ce mécanisme est également valable chez d'autres espèces, mammifères ou non. (Bi, 2012).

b-L'hétérochromatine facultative

Cette hétérochromatine inclut des régions d'ADN codant du génome au sein desquelles la répression transcriptionnelle des gènes dépend de décisions développementales spécifiques. Selon le type cellulaire, telle ou telle portion du génome peut être condensée, et de ce fait inactivée. De la même façon, pour un type cellulaire donné, une même portion de chromatine peut être plus ou moins condensée, selon le niveau de différenciation. Il existe dans la littérature un exemple phare d'hétérochromatinisation facultative : l'inactivation du X chez les femelles mammifères.

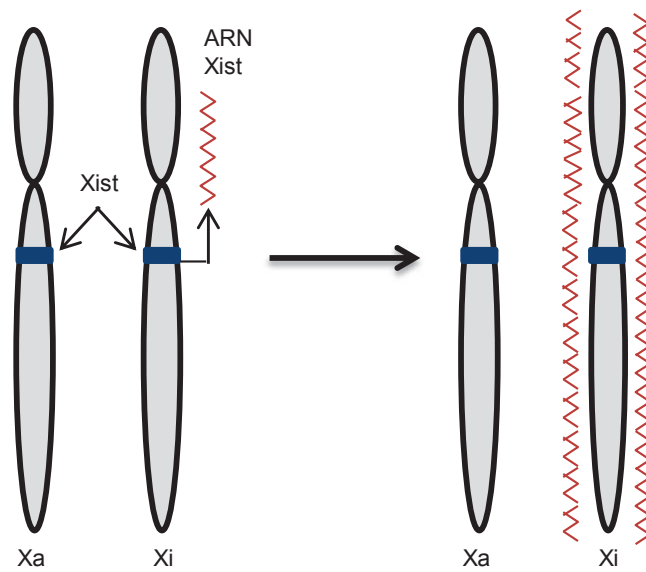


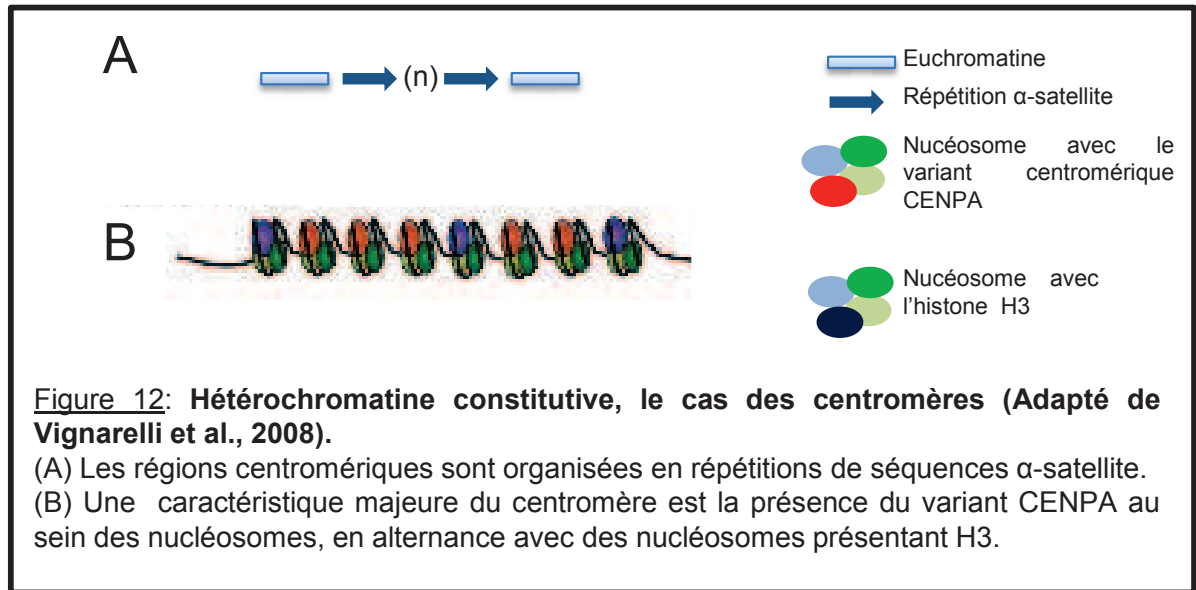
Figure 11: Inactivation du X chez la cellule mammifère femelle.

Chez les mammifères, le mâle possède deux chromosomes sexuels distincts, un X et un Y alors que la femelle présente deux X. Afin de compenser ce déséquilibre, un mécanisme d'inhibition d'un X (Xi) dans les cellules femelles a été mis en place, impliquant l'expression d'un ARN codant de grande taille, Xist qui recouvre le Xi et empêche ainsi l'expression des gènes de ce chromosome.

c-Cas du Xi

Chez les mammifères, les mâles possèdent une copie de chaque chromosome sexuel, un X et un Y alors que les femelles présentent deux X. Dans le but d'éviter un déséquilibre fonctionnel dû à une surexpression génique, les mammifères ont développé un mécanisme de compensation consistant en l'inactivation d'un des deux chromosomes X dans les cellules femelles. Ce phénomène a lieu très tôt lors de l'embryogenèse, lorsque les cellules engagent le processus de différenciation à l'origine des 3 lignages primordiaux de l'embryon : l'ectoderme, l'endoderme et le mésoderme. Il aboutit à l'extinction aléatoire de la transcription de l'un des deux X dans chaque cellule. Le Xi montre par microscopie des propriétés caractéristiques de l'hétérochromatine constitutive: il reste visible dans les cellules au cours de l'interphase sous la forme d'un corpuscule de Barr et il est répliqué tardivement au cours de la phase S. Il est cependant important de mettre en avant que ce Xi fait bien partie de l'hétérochromatine facultative et non pas constitutive, avec entre autres une absence de séquences satellites répétées, caractéristiques de l'hétérochromatine constitutive.

L'inactivation du X fait intervenir dans un premier temps une étape de « sensing », au cours de laquelle chaque cellule détermine le nombre de X présent par rapport au nombre d'autosomes. Cette étape répond à une règle stricte, celle du n-1 découverte en 1967 par Ohno stipulant que quelque soit le nombre de X dans la cellule, tous seront inactivés sauf un. Les étapes d'inactivation du X sont contrôlées par un locus présent sur ce chromosome, appelé centre d'inactivation ou Xic. Ce locus contient entre autres le gène Xist qui est primordial pour ce processus d'inactivation (Brown *et al.*, 1991). En effet, son ARN non codant de grande taille a la capacité à recouvrir le X qui le produit, induisant ainsi son extinction (Heard et Distèche, 2006) (**Figure 11**). L'expression de Xist est contrôlée par un second ARN localisé dans la même région mais issu d'une transcription anti sens : Tsix (Lee *et al.*, 1999). Une hypothèse émise par Rastan en 1983 propose que le X choisi pour rester actif serait le seul où l'expression de l'allèle Xist est bloqué, la nature du facteur bloquant restant encore inconnue. Le X inactivé est également soumis à des modifications épigénétiques telles que (i) des modifications post traductionnelles des queues d'histones comme la triméthylation des H3K9 et H3K27 (H3K9me3 et H3K27me3) trouvées de façon régionales sur le Xi (Chadwick et Willard, 2005) mais également de la H4K20 (H4K20me3), une marque préférentiellement associée à l'hétérochromatine constitutive, la déacétylation des histones H2A, H2B, H3 et H4 (O'Neill *et al.*, 2003); (ii) l'inclusion de variants d'histones tel que Macro H2A (Costanzi et



Perhson, 1998); et (iii) la liaison de protéines caractéristiques de l'hétérochromatine comme HP1. La combinaison des modifications d'histones, de la présence de variant et de l'incorporation de protéines non histone en plus de Xist serait à l'origine de la formation d'un état transcriptionnel silencieux du X ciblé ainsi que de sa réplication tardive et de son apparence condensée dans le noyau.

Il est important de noter cependant que certains gènes (environ 15%) présents au niveau du X inactif ont la capacité à s'échapper de ce silence transcriptionnel. Ces gènes sont de deux sortes. Les premiers sont des gènes également présents au niveau du chromosome Y, le phénomène de compensation de dosage n'étant donc pas requis. Les seconds sont des gènes présents au niveau du bras court du chromosome X, acquis récemment au cours de l'évolution. Illustrant également le caractère principalement facultatif de l'hétérochromatinisation du Xi, celui-ci est réactivé dans la lignée germinale à l'origine des gamètes. Ceci a lieu en même temps qu'une reprogrammation épigénétique générale incluant l'effacement de l'empreinte parentale ainsi qu'une déméthylation de l'ADN sur tout le génome.

d-L'hétérochromatine constitutive

Par définition, ce type de chromatine correspond à des portions de chromosomes apparemment inactives dans toutes les cellules et porteuses de séquences répétitives. La plus grande partie de ces régions se trouve à proximité des centromères (régions péri-centromériques) et des télomères et est répliquée tardivement au cours de la phase S.

Le cas des centromères : les centromères sont des régions spécialisées des chromosomes eucaryotes qui permettent la ségrégation des deux chromatides sœurs de façon équitable entre les deux cellules filles lors de la division cellulaire. Les centromères sont constitués de répétitions de séquences satellites particulières, dont le nombre est variable selon les espèces (**Figure 12A**). L'introduction d'un plasmide codant ce type de séquences satellites au sein d'une cellule est suffisant à lui seul à la création de néo-centromères (Harrington *et al.*, 1997). Une seconde caractéristique des centromères est la présence du variant histone : CENP-A présent dans les nucléosomes, en alternance avec l'histone H3 (Blower *et al.*, 2000) (**Figure 12B**). Les modifications post-traductionnelles des histones présents au niveau de cette hétérochromatine constitutive sont principalement un fort taux de H3K9me3 et une faible acétylation des histones. On peut également noter la présence de la protéine HP1. Paradoxalement, il est possible de trouver des histones diméthylés sur leur lysine 4

A



B

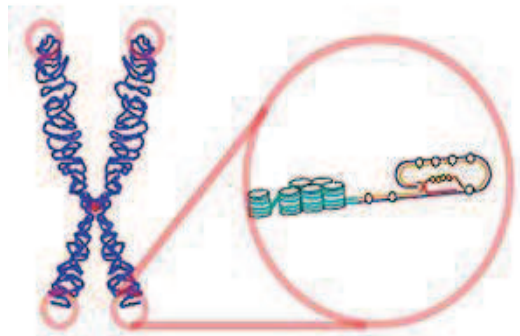


Figure 13: **Hétérochromatine constitutive, le cas des télomères (mercola.com).**

(A) La fin des chromosomes eucaryotes présente une structure spécifique appelée télomère, constitués de répétitions de séquences « TTAGGG ». A l'exception des cellules indifférenciées, la taille de ces structures diminue à chaque division cellulaire du fait de l'action de l'ADN polymérase et de l'absence de télomérase.

(B) Les structures mêmes des télomères permettent une protection contre des fusions entre chromosomes et protègent les extrémités des chromosomes contre l'action d'exonucléases.

(H3K4me2), une marque plus rattachée à l'activité transcriptionnelle. D'autre part, dans les noyaux interphasiques, les centromères de plusieurs chromosomes seraient capables de se regrouper afin de former des chromocentres visibles dans tous les types cellulaires (Alcobia et Dilao, 2000).

Le cas des télomères : La fin des chromosomes eucaryotes sont des structures spécifiques nommées télomères. Ils sont constitués de une à plusieurs dizaines de répétitions d'une séquence de type «TTAGGG» sur plusieurs kilobases (**Figure 13A**). Les télomères, du fait de leur structures sont des éléments protecteurs des chromosomes vis à vis de possibles fusions entre leurs extrémités, mais également de l'intervention de la machinerie de réparation de l'ADN (qui considère la fin des chromosomes comme des cassures doubles brins) ou bien de leur dégradation par des exonucléases (**Figure 13B**). Dans les cellules encore indifférenciés, la taille des télomères qui diminue à chaque réplication est restaurée par l'action de la télomérase. En effet, à chaque division cellulaire, cette structure est raccourcie du fait du mécanisme d'action de l'ADN polymérase qui nécessite la présence d'une extrémité 5' pour pouvoir répliquer le brin d'ADN. Le brin 3'-OH est donc répliqué à l'aide de petits fragments d'ARN, appelé fragments d'Okasaki qui, pour pouvoir jouer leur rôle, ont besoin de s'hybrider à l'extrémité 3'-OH, hybridation qui se fait à une certaine distance de l'extrémité entraînant ainsi un raccourcissement à chaque division cellulaire. Au-delà d'une certaine limite, ce raccourcissement finit par entraîner une sénescence répllicative c'est-à-dire une mort cellulaire due à la réplication (Blasco, 2005). Les régions télomériques présentent des propriétés chromatiniennes similaires à celles des régions centromériques. De plus des études portant sur les modifications épigénétiques des télomères chez l'Humain et la Drosophile ont montré que ces régions mobilisaient HP1 et présentaient des H3K9 méthylées qui seraient nécessaires à l'élongation des télomères (Perrini *et al.*, 2004)

2-L'euchromatine

L'euchromatine est de la chromatine décondensée en une fibre de 11nm, riche en gènes (**Figure 10**). Elle est organisée de façon différentielle dans le temps et dans l'espace en fonction de l'état transcriptionnel. Les histones H3 présentes au sein de cette chromatine sont acétylées de façon importante et hypométhylées. Elle inclut chez l'Homme plusieurs types de séquences, à savoir :

- des gènes dont le nombre est variable selon le contexte (type cellulaire, signaux environnementaux ...)

- de l'ADN non codant répété
- de l'ADN non codant non répété
- de l'ADN codant dupliqué
- de séquences non annotées

3-Les grands domaines de chromatine

Il apparaît maintenant, avec l'émergence de techniques d'analyse à grande échelle, que cette vue de la chromatine opposant hétérochromatine et euchromatine nécessite une classification plus fine en domaines fonctionnels de la chromatine, dont le nombre varie en fonction de l'organisme étudié.

Cas de *C. elegans* : Les chromosomes de cet organisme sont différents de ceux étudiés chez les mammifères: il s'agit de chromosomes holocentriques, qui peuvent être divisés en domaines : un domaine central et deux bras présentant beaucoup de répétitions de séquences et un appauvrissement important en gènes. Au cours du projet ENCODE, lancé en 2003, chargé d'étudier le génome Humain, le modèle *C.elegans* a été intégré (2007). L'étude de 19 marques histones, de 2 variants d'histones (H2A.Z et H3.3) et de plusieurs types de protéines associées à la chromatine sur cet organisme modèle a permis de définir 3 groupes principaux de chromatine, pas forcément homogènes dans la nature des séquences qu'ils incluent, mais montrant un enrichissement spécifique en marques histones et en protéines (Gerstein *et al.*, 2010).

- Groupe 1 : il présente un ensemble de marques regroupant la méthylation de H3K4 et H3K36, l'acétylation de résidus d'H3 et une déplétion en H3K9me1/me2 et me3. Ce groupe semble enrichi dans la région centrale des autosomes.
- Groupe 2 : il contient des H3 mono, di ou tri méthylés sur leur K9 et est présent sur les bras des autosomes. Malgré la présence de ces marques généralement associées à une chromatine condensée, le marquage au DAPI ne montre pas d'hétérochromatine sur ces régions qui sont pourtant fortement associées à la lamina par l'intermédiaire de la protéine LEM-2.
- Groupe 3 : il est associé au chromosome X qui est très différent des autosomes. On y trouve différentes marques histones, uniformément distribuées, telles que H4K20me1, H3K9me1 et H3K27me1, souvent associées aux gènes fortement exprimés sur ce

chromosome, ainsi qu'une interaction faible avec LEM-2. Les chromosomes X des hermaphrodites (XX) sont tous les deux soumis à une compensation de dosage et ceci par le recrutement des protéines DPY-26, 27 et 28 et de SDC-3.

Cas de *D. melanogaster* : Dans cette espèce, il existerait au minimum deux types d'hétérochromatine différents, marqués par des protéines différentes. Le premier, marqué par les protéines polycomb ainsi que la méthylation d'H3K27. Ce type serait impliqué dans la régulation des gènes jouant un rôle dans le développement (Sparmann et van Lohuizen, 2006). Le second type est caractérisé par la présence d'HP1 et de protéines associées, ainsi que par la méthylation de l'H3K9. On retrouve cette hétérochromatine en particulier aux alentours des centromères. Ces deux formes distinctes d'hétérochromatine ne montrent aucun chevauchement dans le génome de *D. melanogaster* (de Wit *et al.*, 2007). Les études menées par Filion et ses collègues (Filion *et al.*, 2010) ont depuis mis en avant l'existence de 5 grands domaines chromatinien dans la lignée embryonnaire Kc167, ayant des caractéristiques propres, comme entre autres l'activité de transcription, leur réplication au cours de la phase S, et les modifications post-traductionnelles des histones présents. Ces domaines ont été nommés vert, bleu, noir, jaune et rouge. Les deux domaines d'hétérochromatine précédemment évoqués ont bien été retrouvés. Tout d'abord, le domaine VERT où l'on note la présence de SU(VAR)3-9, HP1, LHR et HP6 ainsi que de H3K9me3. Ce domaine est principalement localisé au niveau des régions centromériques et péri-centromériques (c'est-à-dire dans l'hétérochromatine constitutionnelle). Ensuite, le domaine BLEU caractérisé par la présence de protéines du groupe polycomb telles que PC, E(Z), PCL et SCE ainsi que par l'enrichissement en H3K27me3. Des gènes connus pour être liés par les protéines polycomb, et impliqués dans le développement comme les gènes HOX ont été identifiés au sein de ce domaine. Cependant, un troisième domaine chromatinien d'environnement répressif a été identifié par Filion *et al.*, (2010), le domaine NOIR, qui couvre 48% du génome de la Drosophile et en fait donc le plus abondant. Ce domaine présente une activité transcriptionnelle quasi inexistante et est caractérisé par la présence d'une combinaison de protéines regroupant H1, D1, IAL et SUUR ainsi que SU(HW), LAM et EFF (fréquemment présent). Sa caractéristique principale est sa capacité à inhiber activement la transcription comme en témoigne les expériences d'insertions de transgènes dans ce type de région. Les deux derniers domaines identifiés (JAUNE et ROUGE), reliés à l'euchromatine, présentent de réelles similitudes comme la présence de la RNA-Pol II, un haut niveau de H3K4me2 et de H3K79me3, un faible enrichissement en H3K9me2 et H3K27me3 ainsi que la présence de

différentes protéines telles que des HDACs, identifiées récemment dans de la chromatine active au niveau de cellules humaines (Wang *et al.*, 2009). Cependant, malgré un certain nombre de points communs, ces deux domaines montrent de grosses différences principalement dans leurs enrichissements pour certains types de facteurs de transcription, mais également dans leur réplication. En effet, même si ces deux types de chromatine sont répliqués relativement tôt au cours de la phase S, le domaine ROUGE est plus précoce que le domaine JAUNE. De plus, ces deux domaines contiennent des gènes impliqués dans des processus biologiques différents avec pour la chromatine JAUNE des processus universels comme la réparation de l'ADN ou bien le métabolisme des acides nucléiques, tandis que les gènes présents dans la chromatine ROUGE seraient plus reliés à des phénomènes tissus spécifiques.

Cas de *H. sapiens*: Les caractéristiques de la chromatine structurant le génome Humain ont particulièrement été étudiées au sein de deux travaux majeurs. La première étude menée par Ernst et Kellis (Ernst et Kellis, 2010) sur des lymphocytes T4 a porté sur l'analyse de 38 modifications post-traductionnelles d'histones (méthylation ou acétylation) et sur l'enrichissement en H2A.Z, RNA-Pol II et CTCF. La seconde a analysé 9 marques post-traductionnelles d'histones dans 9 types cellulaires différents (Ernst *et al.*, 2011). Ces deux travaux ont permis de déterminer 6 classes principales de chromatines:

- les promoteurs, enrichis en îlots CpGs et en motifs conservés pour la liaison de différents facteurs de transcription, divisibles en 3 sous groupes:
 - * actifs, présentant un enrichissement important en RNA-Pol II et en H2A.Z, un fort taux de méthylation de H3K4 ainsi que d'acétylation de H3K27 et de H3K9 et un enrichissement en sites sensibles à la DNase I
 - * activables, caractérisés par des H3K4 méthylés et des H3K27 non acétylés
 - * inactifs, montrant un faible taux de méthylation de l'H3K4 mais un fort enrichissement en H3K27me3
- les enhancers, de deux types:
 - * actifs, marqués par des H3K4 méthylés, de nombreuses acétylations (H2BK120, H3K27 et H2BK5) ainsi qu'une importante sensibilité à la DNase I et un enrichissement fort en H2A.Z



Figure 14: Organisation des territoires chromosomiques d'un noyau eucaryote (Bolzer et al., 2005).

Dans les noyaux interphasiques des cellules eucaryotes, les chromosomes ne sont pas distribués au hasard mais occupent des régions précises. Les chromosomes pauvres en gènes seraient situés en périphérie nucléaire et les chromosomes riches en gènes à l'intérieur du noyau.

* inactifs, ne présentant pas d'acétylation ainsi qu'une sensibilité à la DNase I plus faible que les enhancers actifs

- les insulateurs, principalement caractérisés par un enrichissement très important de CTCF
- les régions transcrites, définie par une combinaison en 7 marques (H3K79me1/me2/me3, H3K27me1, H2BK5me1, H4K20me1 et H3K36me3). La chromatine transcrite peut être divisée en 3 sous groupes:

* les exons épissés, marqués par une combinaison spécifique de H3K36me3, H2BK5me1, H4K20me1, H3K79me3 et l'absence de H3K4me2, H3K9me1 et H3K79me2/me3

* les sites de fin de transcription (TES, transcriptional end site), enrichis en H3K36me3, H4K20me1 et en RNA-PII mais ne présentant pas de H3K4me1/me2/me3

* une catégorie spécifique représentant les gènes à doigts de zinc, qui sont fortement enrichis en H3K9me3, H4K20me3, H3K36me3 et appauvris pour les autres modifications post-traductionnelles citées précédemment

- les régions non transcrites, mobilisant des protéines de type polycomb et souvent associées à H3K27me3
- l'hétérochromatine et ses régions répétées, qui diffèrent des régions non transcrites par leur enrichissement en H3K9me3

D-Niveaux supérieurs de l'organisation chromatinienne

1-Notion de territoires chromosomiques

La notion de territoires chromosomiques (CTs) a été suggérée dès 1885 par Rabl et Boveri mais n'a été confirmée par FISH (fluorescent *in situ* hybridization) qu'un siècle plus tard. Il existe en effet au sein du noyau interphasique une organisation précise des chromosomes, permettant une ségrégation spatiale du matériel génétique de chaque chromosome non pas au hasard mais dans des régions définies (Misteli, 2005) (**Figure 14**). Ces territoires sont des structures non solides. Les chromosomes présentant le plus grand nombre de gènes seraient présents vers l'intérieur du noyau, au contraire des chromosomes les

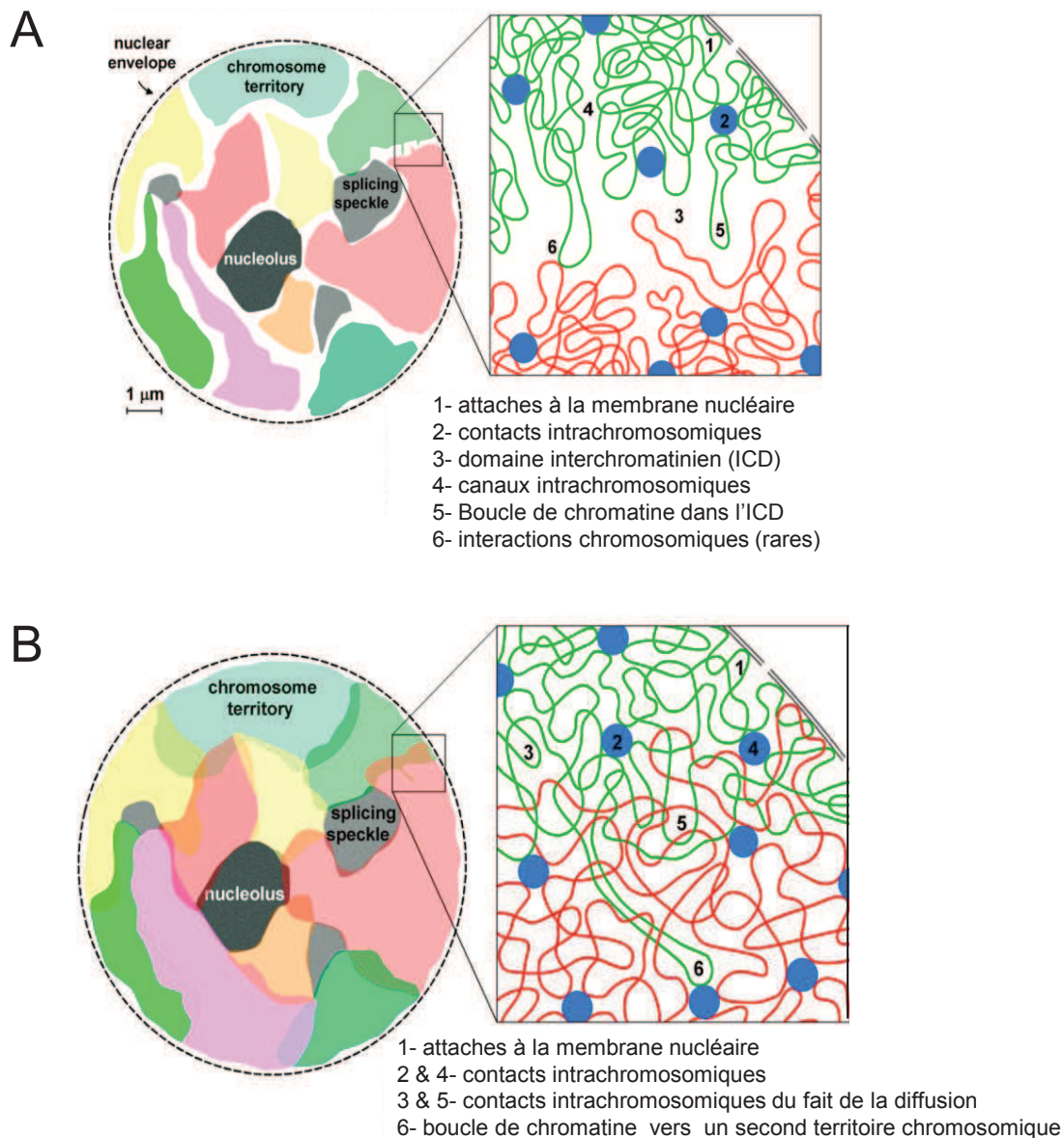


Figure 15: Modèles d'organisation des territoires chromosomiques dans les cellules mammifères (Branco et Pombo, 2006).

(A) Dans le modèle des domaines interchromatiniens, les différents territoires chromosomiques (CTs) sont séparés par un compartiment interchromatinien riche en protéines appartenant à la machinerie nucléaire. Dans ce modèle, les gènes actifs sont présents en périphérie de CTs et les boucles de chromatine sont rares.

(B) Dans le modèle du réseau interchromosomique (ICN), les chromosomes ne sont pas séparés par un compartiment mais sont entremêlés.

plus pauvres en gènes sont principalement localisés en périphérie nucléaire (Croft *et al.*, 1999 ; Cremer *et al.*, 2001). Cette généralité est cependant à nuancer d'une part du fait d'une organisation en « patchwork » des chromosomes avec des régions plus ou moins denses en gènes impliquant donc une localisation périphérique de certains chromosomes riches en gènes (Mahy, 2002) et d'autre part du fait d'une organisation variable selon les types cellulaires avec par exemple une organisation en fonction de la densité génique (plus que l'activité génique individuelle) dans les lymphocytes ou bien une organisation en fonction de la taille des chromosomes dans les fibroblastes (Misteli, 2005). Actuellement, il existe deux modèles pour l'organisation des CTs:

Le modèle des domaines inter-chromatiniens (ICD): dans ce modèle, tel qu'il a été proposé en 1993 par Zirbel *et al.*, des transcrits de gènes et des composants de la machinerie d'épissage des ARNs sont concentrés tout le long des CTs. Cependant, cette notion de CTs très compacte impliquerait que la transcription et l'épissage aient lieu uniquement en surface des CTs et donc que les gènes transcrits seraient situés en périphérie de ces domaines tandis que les séquences non codantes seraient à l'intérieur. Or, un certain nombre d'études ont montré une distribution équilibrée des domaines riches et pauvres en gènes le long des chromosomes (Visser *et al.*, 1998 ; Mahy *et al.*, 2002b). Il existe donc un modèle ICD légèrement modifié dans lequel les CTs sont traversés d'autres structures, des canaux, ayant pour but de faciliter l'accès à des facteurs de régulation. La transcription ainsi que l'épissage des ARNs se ferait dans cet espace entre les chromosomes (Cremer et Cremer, 2001) (**Figure 15A**).

Le modèle du réseau interchromosomal (ICN): contrairement au premier modèle, celui-ci, proposé par le laboratoire d'Ana Pombo (Branco et Pombo, 2006), estime qu'il n'y a pas de séparation physique des CTs comme en témoignent les possibilités d'emmêlement des différents CTs voisins ainsi que leur capacité de translocation inter-chromosomique. Seule la présence des CTs adjacents, de la membrane nucléaire et des différents compartiments nucléaires cités précédemment pourrait délimiter les capacités d'emmêlement (**Figure 15B**).

2-Protéines impliquées dans l'organisation des boucles de chromatine

a-La cohésine

La cohésine a été décrite pour la première fois en 1997 par Michaelis *et al.*, Cette protéine est composée de 4 sous unités protéiques conservées au cours de l'évolution

Tableau 1: Sous unités de la cohésine dans diverses espèces modèles (Adapté de Onn et al., 2008).

	<i>S.cerevisiae</i>	<i>S.pombe</i>	<i>D.melanogaster</i>	<i>X.laevis</i>	<i>H.sapiens</i>
Sous unité de la cohésine	SMC 1 SMC 3 SCC 1 SCC 3	Psm 1 Psm 3 Rad 21 Psc 3	Smc 1 Smc 3 Rad 21 SA	Smc 1 Smc 3 Rad 21 SA 1, SA 2	Smc 1 Smc 3 Rad 21 (Scc 1) SA 1 (STAG 1) SA 2 (STAG 2)

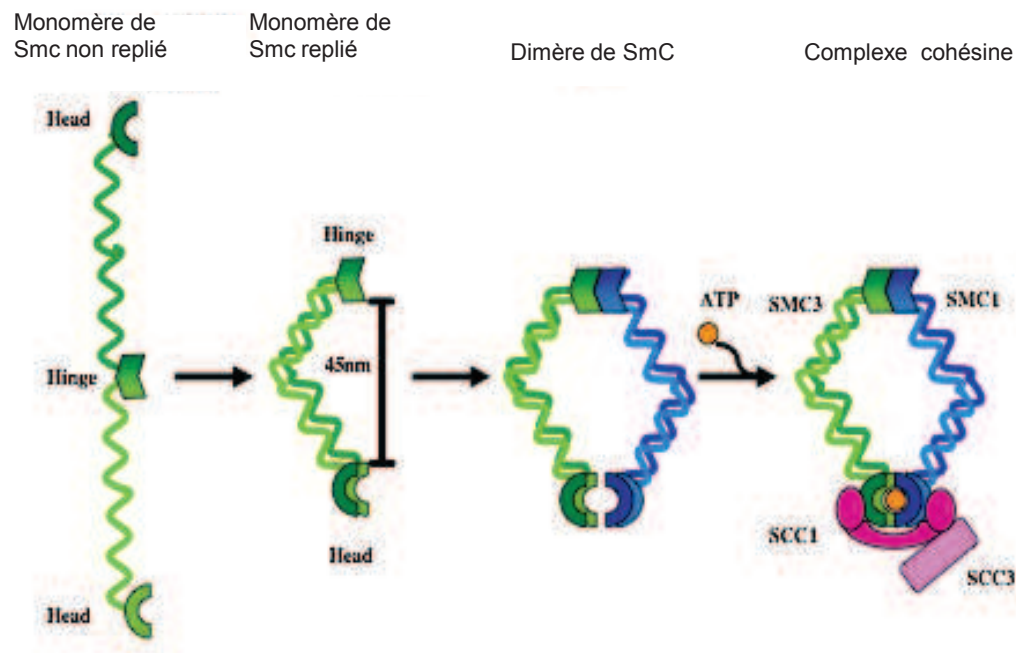


Figure 16: Sous-unités et structure de la cohésine (Adapté de Onn et al., 2008).

L'étape initiale de la formation du complexe cohésine est le repliement de la protéine Smc qui permet la formation d'un domaine « hinge » et d'un domaine « head », suivi d'une homodimérisation avec une seconde protéine Smc repliée. La liaison de SCC1, sous-unité à activité ATP-dépendante, est la troisième étape et permet le recrutement de SCC3, ce qui aboutit à terme à la formation du complexe de la cohésine.

(**Tableau 1**), SCC1 et SCC3 ainsi que SMC1 et SMC3 qui sont impliquées dans le maintien structurel des chromosomes. Des homologues de ces sous-unités ont été identifiés de la bactérie à l'Homme (Hirano, 2005). Les protéines SMC sont des protéines très conservées à la fois dans leurs séquences et dans leurs structures et dotées d'une activité ATPase. Elles ont la capacité à se dimériser *via* leur domaine charnière (« hinge domain »), et sont ensuite connectées par SCC1 qui lie également SCC3 pour former le complexe cohésine (**Figure 16**). La cohésine a également la capacité à s'associer avec un grand nombre de protéine partenaires. Des défauts dans les sous unités SMC1 et MSC3 ou bien dans la protéine permettant de « charger » la cohésine sur l'ADN (Nipbl) sont à l'origine du syndrome Cornelia de Lange (Liu et Krantz, 2009), caractérisé par des problèmes dans le développement tels que des retards mentaux, des retards de croissance, l'autisme...

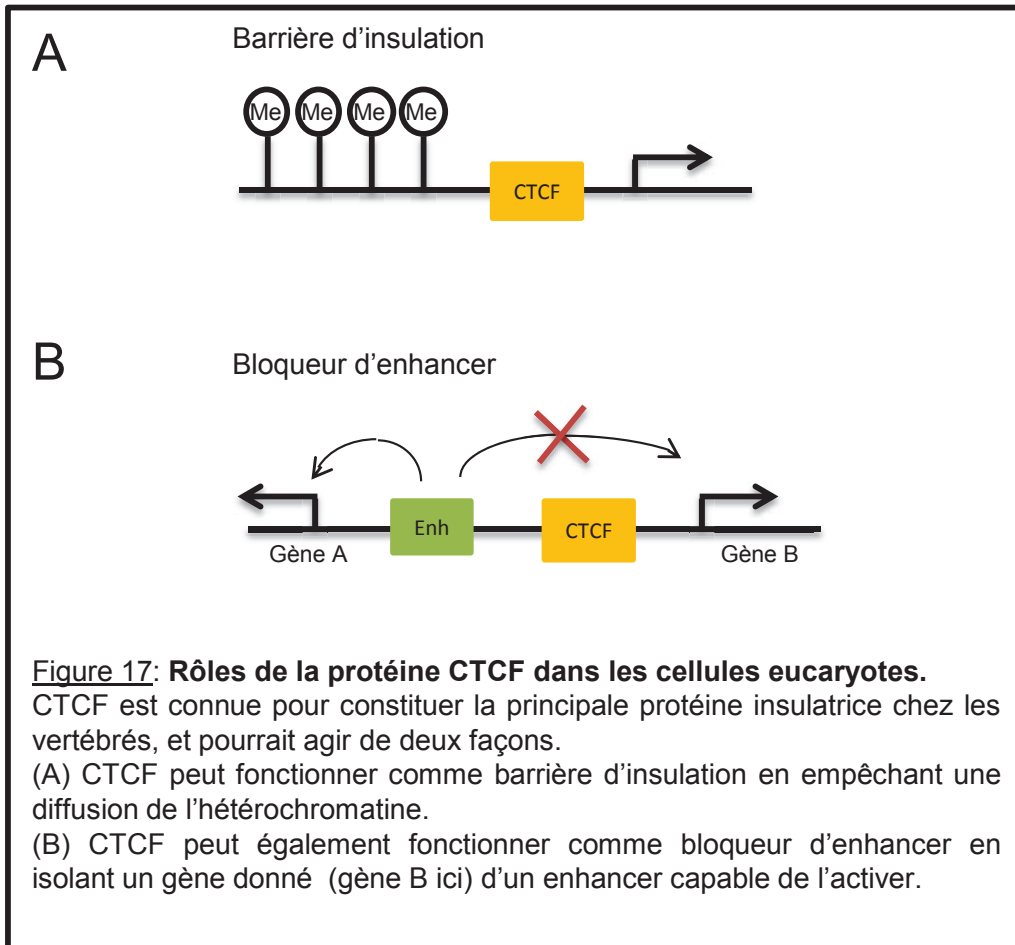
Le rôle premier de la cohésine est de prévenir une séparation précoce des chromatides sœurs lors de la division cellulaire afin de garantir une prolifération cellulaire normale (Wendt et Peters, 2009). Cependant, l'identification, chez les vertébrés, d'une liaison de la cohésine à la fin de la mitose, bien avant la nécessité d'une cohésion des chromatides pour le cycle cellulaire a suggéré une possible fonction de ce complexe sur l'ADN en dehors de la réplication (Wendt *et al.*, 2008). Des expériences de ChIP-chip dans des cellules post-mitotiques ont ainsi permis de montrer une liaison de la cohésine à la chromatine indépendante de son rôle dans la cohésion (Wendt *et al.*, 2008) et principalement intergénique (49% contre 35% dans les introns et 13% 5kb autour de gènes). En ce qui concerne la régulation des gènes, le rôle de la cohésine semble plus ou moins complexe selon les organismes étudiés. Chez la levure *S.pombe*, elle permet de mettre fin à la transcription entre deux gènes convergents et cela à travers l'activité de DICER (une protéine connue pour son implication dans la machinerie des miRNAs) et de Swi6. Lors de ce processus, la cohésine est recrutée avant et pendant la phase S et sa présence en phase G2 bloque l'action de la RNA-Pol II dans les régions intergéniques (Gullerova et Proudfoot, 2008). Chez la Drosophile, des travaux effectués sur l'insulateur gypsy ont permis de mettre en avant le rôle de l'orthologue de Scc2, Nipped B, dans l'interaction entre les régions promotrices et enhancers des gènes cut et ultrabithorax (Rollins *et al.*, 1999). La cohésine permettant de garder une proximité importante entre les deux chromatides, ce complexe possède également un rôle dans la réparation des cassures doubles brins, la chromatide sœur servant de modèle dans la réparation de la chromatide cassée. Il est également envisageable qu'au cours de ce processus

de réparation, la cohésine permette de garder une intégrité physique à la chromatide elle-même (Wendt et Peters, 2009).

b-CTCF

CTCF est une protéine de liaison à l'ADN très conservée chez les eucaryotes supérieurs, notamment au niveau de son domaine central incluant 11 doigts de Zinc (Phillips et Corces, 2009). Ces derniers sont capables de se combiner différenciellement selon la séquence cible dans le but de médier au mieux la liaison à la chromatine (Filippova *et al.*, 1996). Ce domaine central est très conservé comme en atteste une homologie de séquence proche des 100% entre la Souris, le Poulet et l'Homme (Phillips et Corves, 2009), les différences portant uniquement sur les extrémités N-terminales et C-terminales. CTCF a été identifiée deux fois de façon totalement indépendante, tout d'abord comme étant une protéine liant des séquences dans la région promotrice de l'oncogène MYC (Lobanenko *et al.*, 1990) mais également comme une protéine liant un élément silencier dans le promoteur du gène de l'amyloïde H chez le Poulet ainsi qu'une protéine se liant au niveau de séquences régulatrices du gène de la β -globine du Poulet (Ohlson *et al.*, 2001). CTCF se lie également sur le chromosome X (Chao *et al.*, 2002), sur les gènes soumis à l'empreinte parentale (Bell et Felsenfeld, 2000) ainsi qu'aux niveau des frontières entre la chromatine active et la chromatine inactive (Cho *et al.*, 2005).

Des expériences de ChIP-chip ont permis de déterminer la distribution des sites de liaison de CTCF sur l'entièreté du génome de fibroblastes humains, les cellules IMR90 (Kim *et al.*, 2007). Ainsi, CTCF se lierait principalement au niveau de régions intergéniques: 46% ainsi qu'au niveau des introns (22%), des exons (12%) et dans une fenêtre de 2 kb autour des promoteurs (20%). Cependant d'autres études plus récentes, utilisant la technique de ChIP-seq ainsi qu'un autre type cellulaire ont montrés une distribution légèrement différente (Barsky *et al.*, 2007) avec 45% de cibles intergéniques, 29% dans les introns, 3% dans les exons, 7% dans la 5'-UTR, 2% dans la 3'-UTR et enfin 13% autour du TSS. Malgré ces différences, 70% des sites de liaison de CTCF semblent être conservés entre les différents types (Kim *et al.*, 2007). Ces observations sont à l'origine de l'hypothèse que les sites de liaison de CTCF sont principalement invariants d'un type cellulaire à l'autre, les différences relativement mineures pouvant être observées illustrant soit une significativité fonctionnelle importante ou bien uniquement un biais de sensibilité dû à l'utilisation de techniques différentes.



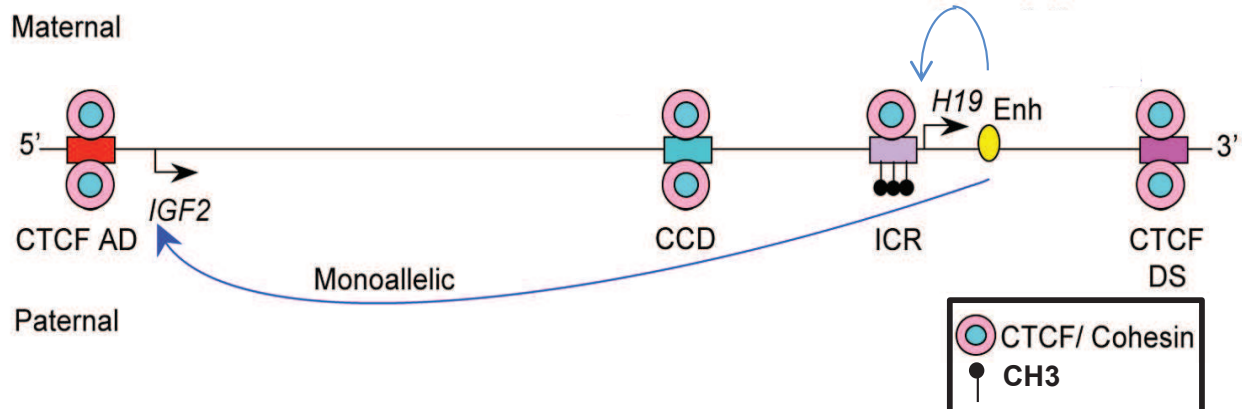
CTCF est considéré comme étant la principale protéine insulatrice chez les vertébrés, pouvant agir soit en tant que bloqueur d'enhancer soit comme barrière d'insulation en prévenant une expansion d'eu- ou d'hétéro-chromatine aux alentours de ses sites de fixation (Wendt et Peters, 2009) (**Figure 17**). En conséquence, CTCF peut exercer une influence directe ou indirecte sur la régulation transcriptionnelle, sa déplétion par transfection de siRNA pouvant entraîner une dérégulation de l'expression de centaines de gènes (Wan *et al.*, 2008). Si les rôles de CTCF varient selon son site de liaison (Phillips et Corces, 2009), son action dépend également de ses partenaires protéiques. En effet, en interaction avec YBI, YYI, Kaiso et Sin 3, CTCF possède plutôt une influence sur la régulation transcriptionnelle de certains gènes, tandis qu'en interaction avec la nucléophosmine et ou CHD8, il aurait plus une fonction d'insulateur (Rubio *et al.*, 2008). Par ailleurs, un ensemble de données suggère que la fonction insulatrice de CTCF repose sur sa capacité à former des boucles de chromatine entre deux séquences distantes sur l'ADN. De plus, l'existence de ces boucles de chromatine CTCF-dépendantes pourrait expliquer la différence observée entre le nombre potentiel de sites de liaison de CTCF prévus par des méthodes computationnelles et le nombres de sites réels identifiés par ChIP (méthode ne permettant pas de distinguer des liaisons directes ou indirectes de CTCF) (Gerasimova *et al.*, 2000 ; Pant *et al.*, 2004 ; Yusufzai *et al.*, 2004). Du fait de ses 11 doigts de Zinc et de leur plasticité dans la liaison à l'ADN, plusieurs «classes» de boucles de chromatine CTCF-dépendante ont été suggérée.

Il semblerait qu'une partie des loci liés par CTCF le resterait à travers la mitose et cela dans un but de protection contre des méthylations ectopiques. Cela permettrait de ré-établir les structures initiales dès la sortie de mitose, structures essentielles pour une conservation du programme transcriptionnel et donc de son phénotype spécifique.

c-Complexe cohésine et CTCF : actions coordonnées et distinctes

CTCF et la cohésine présentent une colocalisation plus ou moins importante selon les types cellulaires (Rubio *et al.*, 2008 ; Wendt *et al.*, 2008), suggérant l'existence d'un interrelation fonctionnelle entre ces deux facteurs. Une hypothèse avancée dès 2008 (Rubio *et al.*, 2008 ; Wendt *et al.*, 2008) est que CTCF serait nécessaire à la liaison de la majeure partie du pool de cohésine au niveau de certains sites spécifiques de l'ADN. Cependant, les formes solubles de ces deux facteurs ne sont pas liées de façon stable dans des extraits nucléaires de cellules HeLa (Wendt *et al.*, 2008). Ceci semblerait exclure l'existence d'une relation directe entre CTCF et la cohésine. L'action de CTCF pourrait être indirecte, en influant sur la

A



B

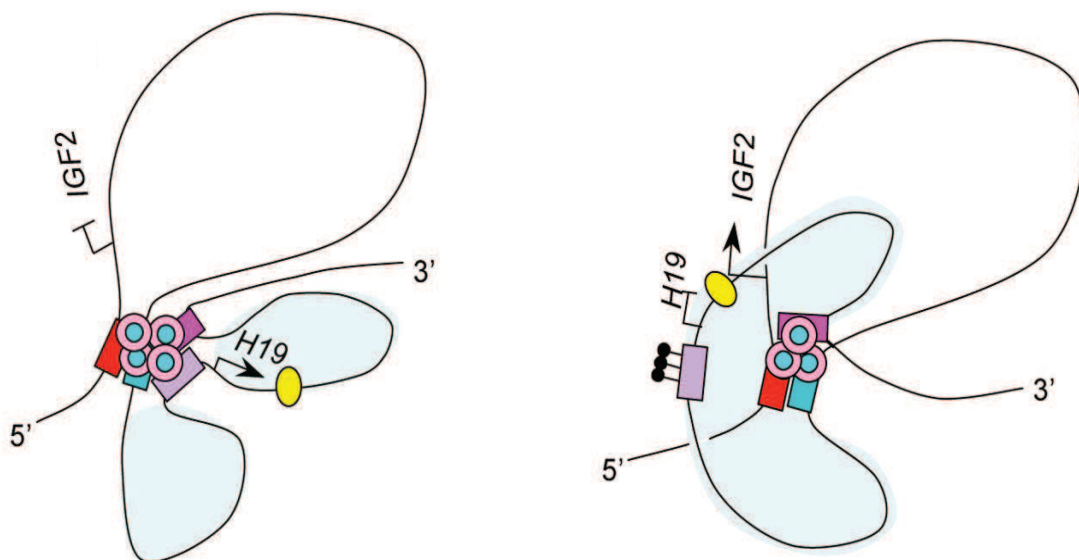


Figure 18: Interactions médiées par CTCF et la cohésine sur le locus *H19/IGF2* (Nativio et al., 2008).

(A) Représentation du locus *H19/IGF2* avec les sites de liaison du complexe cohésine/CTCF (proche de *IGF2*: CTCF AD; au niveau d'un site d'hypersensibilité à la DNase I: CCD, au niveau de l'ICR (Imprinting Control Region) et en amont de *H19*: CTCF DS) ainsi que les éléments régulateurs (ICR en violet, l'enhancer en jaune). Au niveau de l'allèle maternel, l'ICR est lié par CTCF ce qui entraîne la formation d'une boucle spécifique mettant l'enhancer à proximité d'*H19* mais loin de *IGF2*. Au contraire, la présence de méthylation sur l'ICR au niveau de l'allèle paternel protège cette séquence régulatrice de la liaison de CTCF, ce qui entraîne la création de boucles de chromatine différentes en relation avec l'activation d'*IGF2* et la répression de *H19*.

(B) Formation des boucles de chromatine par l'intermédiaire du complexe CTCF/cohésine, différentes selon que l'on soit au niveau de l'allèle paternel ou de l'allèle maternel.

structure de la chromatine à un locus donné, entraînant une accumulation de la cohésine à cet endroit (Wendt et Peters, 2009). D'un autre côté, il existe une partie du pool de la cohésine qui a la capacité de se lier à la chromatine sans l'aide de CTCF ; mais *via* l'interaction avec d'autres protéines, la cohésine n'ayant pas la capacité à se lier directement à l'ADN faute de domaine de liaison. Par exemple, de tels sites cohésine sans CTCF ont été identifiés à proximité de sites de liaison pour des facteurs de transcription spécifiques comme ER α dans les MCF-7 ou bien HNF4 α dans les cellules HepG2 (Schmidt *et al.*, 2010).

Le locus soumis à l'empreinte parentale, IGF2/H19 est un très bon exemple du rôle de ces deux protéines au sein du génome. Une séquence insulatrice, localisée en amont du gène H19, connue sous le nom de «région de contrôle de l'empreinte» (ICR, imprinting control region), permet de contrôler l'expression spécifique du gène H19 par l'allèle maternel et la transcription de IGF2 *via* l'allèle paternel. Cette empreinte parentale est dépendante de l'action de CTCF, qui, lié au niveau de l'ICR de l'allèle maternel, prévient l'action d'un enhancer distal sur IGF2 qui par défaut active H19. Au contraire, sur l'allèle paternel, la liaison de CTCF est bloquée par la présence de CpG méthylés sur l'ICR permettant ainsi à l'enhancer d'activer IGF2 plutôt que H19 (Wendt *et al.*, 2008). CTCF est également nécessaire au recrutement de la cohésine au niveau de ce locus, permettant la formation et la stabilisation de boucles de chromatine essentielles à la régulation des gènes IGF2 et H19 (Nativio *et al.*, 2008) (**Figure 18**).

Il existe d'autres exemples bien décrits de régulations géniques concertées par CTCF et la cohésine *via* la formation de boucles de chromatine, comme les loci codant pour l'interféron gamma (Hadjur *et al.*, 2009) ou codant la β -globine (Hou *et al.*, 2009).

3-Etudes à grande échelle des interactions inter et intra chromosomiques (3C, ChIA PET, HiC)

Bien que les informations génomiques soient présentées le plus souvent comme reposant sur une série linéaire de bases, l'organisation des génomes dans la 3^{ème} dimension par des interactions entre des domaines chromatinien distants portés sur un même chromosome ou des chromosomes distincts est un paramètre critique à considérer lorsque l'on veut étudier des régulations transcriptionnelles. La réalisation d'expériences de ChIP-chip ou plus récemment de ChIP-seq a ainsi par exemple permis de montrer que les sites de liaisons des facteurs de transcription sont le plus souvent localisés loin en amont ou en aval des promoteurs des gènes qu'ils sont susceptibles de réguler (Carroll *et al.*, 2005 ; Massie et

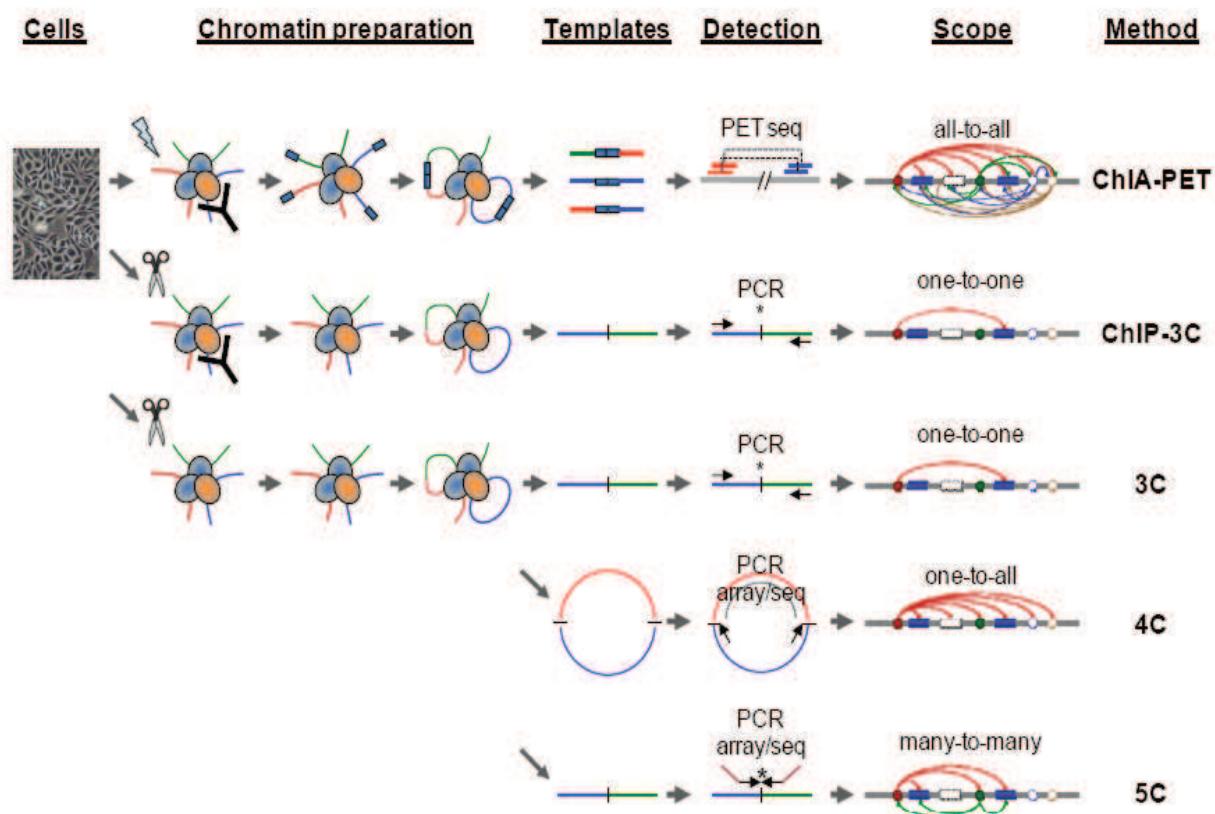


Figure 19: Schéma des techniques d'analyses des interactions inter- et intra-chromosomiques (Fullwood et al., 2009).

Ces cinq techniques commencent par une étape commune de « cross-linking » des cellules au formaldéhyde afin de figer les interactions inter- et intra-chromosomiques, suivie d'une étape de fragmentation de la chromatine par sonication ou *via* une enzyme de restriction. Dans le cas des techniques de 3C, 4C et 5C, les fragments de chromatine sont dilués avant de subir une ligation, permettant la synthèse de fragments d'ADN hybrides entre des régions génomique proches dans l'espace mais lointaines au niveau de la séquence primaire du génome. Les interactions sont analysées par qPCR (3C), sur puces à ADN ou bien par séquençage haut débit dans le cas des techniques de 4C et 5C.

Pour les techniques de ChIP-3C et de ChIA PET, une étape initiale de ChIP est effectuée afin d'enrichir la fraction chromatiniennne analysée en boucles de chromatine impliquant une protéine d'intérêt donnée.

Mills, 2008). C'est la mise au point de la technique de 3C (« chromosome conformation capture ») ainsi que de ses dérivés (**Figure 19**) qui a permis de mettre en évidence la présence de boucles de chromatine capables de réunir à proximité des sites de liaison d'un facteur de transcription donné et les promoteurs de ses gènes cibles (West et Fraser, 2005 ; Woodcock, 2006). Cette technique du 3C, comme le ChIP, est basée sur la fixation au formaldéhyde des liaisons existantes entre les protéines ainsi qu'entre les protéines et l'ADN. La chromatine récupérée est digérée par une enzyme de restriction avant d'être diluée puis soumise à une étape de ligation. Cette étape permet d'obtenir des fragments d'ADN chimériques composés de séquences éloignées au niveau de la séquence primaire du génome mais qui sont de fait proches dans l'espace nucléaire. Ces fragments révélateurs d'interactions intra et inter-chromosomales sont, après réversion de la fixation au formaldéhyde, amplifiés par qPCR (Dekker *et al.*, 2002). Un premier travail réalisé au sein du modèle levure en utilisant cette technique a permis d'acquérir de nombreuses informations sur l'organisation tri-dimensionnelle des chromosomes et du noyau en général (Dekker *et al.*, 2002). La première transposition de la méthode de 3C dans des cellules d'eucaryotes supérieurs, menée sur un modèle murin, a permis de confirmer l'existence de boucles de chromatine permettant de rapprocher des éléments régulateurs distants aux promoteurs de gènes donnés au niveau du locus de la β -globine (Tolhuis *et al.*, 2002).

Cependant cette technique reste limitée car elle oblige à se focaliser sur un site en particulier (même si ce problème a été aboli avec le 4C/5C et le Hi-C) et ensuite car elle ne permet pas d'identifier un réseau de boucles mis en place par l'intermédiaire (direct ou non) d'une protéine d'intérêt. Une nouvelle approche, mise au point récemment (Fullwood *et al.*, 2009a), dénommée le ChIA PET résous en partie les problèmes posés par le 3C dans le sens où elle donne la possibilité d'identifier *de novo* des boucles de chromatine impliquant une protéine d'intérêt. Cette technique se rapproche du couplage d'un ChIP et d'un 3C, avec une première étape de fixation au formaldéhyde suivi d'un ChIP, qui permet d'enrichir la fraction soluble en ADN lié par une protéine d'intérêt. S'ensuit une étape de liaison de «linker» aux extrémités des fragments récupérés après le ChIP, de ligation des fragments proches spatialement et enfin d'un séquençage des fragments obtenus (**Figure 19**). Cette technique permet donc d'appréhender au niveau de l'entière du génome l'ensemble des relations existantes entre les régions génomiques recrutant le facteur d'intérêt (cistrome) et d'autres régions spatialement proches. Le premier ChIA PET effectué a utilisé ER α comme protéine d'intérêt, et a permis d'identifier des boucles intra et inter-chromosomiques, même s'il est important de noter que

ER α semble être plus impliqué dans des relations intra-chromosomiques (Fullwood *et al.*, 2009b). Cette étude a confirmé la présence de sites de liaison d'ER α à proximité et à distance des promoteurs de ses gènes cibles, les sites proximaux étant enrichis en H3K4me3 ainsi qu'en RNA-Pol II tandis que les sites distaux présentent un enrichissement préférentiel en FOXA1, une protéine importante pour la liaison d'ER α . Les sites impliqués dans des interactions longues distances présentent cependant un enrichissement plus fort en RNA-Pol II ou en FOXA1 que les autres sites de liaison d'ER α , ce qui fait dire aux auteurs de l'étude que ces deux protéines peuvent permettre de prédire l'existence d'interactions longues distances (Fullwood *et al.*, 2009b). Comme décrit plus-haut, CTCF est proposé comme étant une protéine organisatrice du génome, capable de coordonner les structures tri-dimensionnelles de la chromatine. Une partie des boucles l'impliquant ont récemment été identifiées par ChIA PET (Handoko *et al.*, 2011), dans les cellules murines ES. Comme pour le ChIA PET ER α , CTCF semble être préférentiellement impliqué dans l'établissement d'un réseau intra-chromosomique que inter-chromosomique. Les résultats de cette étude ont montré que CTCF configure le génome en domaines distincts de la chromatine, présentant une signature chromatinienne unique. En effet, l'alignement de signaux de ChIP-seq pour différentes modifications post-traductionnelles d'histones sur les régions impliquées dans ces boucles de chromatine CTCF-dépendantes a permis de mettre en avant l'existence de 5 groupes:

- catégorie I: elle représente 12% des boucles et est enrichie en H3K4me1 et me2, typique d'une chromatine active ainsi qu'une déplétion de H3K9me, H3K20me2 et H3K27me.
- catégorie II: elle représente 11% des boucles, un enrichissement en H3K9me, H3K20me et H3K27me ainsi qu'une sous représentation des marques actives.
- catégorie III: elle représente 19% des boucles et correspond aux enhancers ainsi qu'aux promoteurs en activité. On y observe la présence de H3K4me1 et me2 au sein des boucles et de H3K4me3 aux frontières. H3K27me3 est quant à lui enrichi à l'extérieur des boucles de chromatine.
- catégorie IV: elle est la plus abondante avec 31% des boucles. L'enrichissement en marques histones n'est pas clairement défini au sein de la boucle en elle-même mais elle présente un enrichissement en H3K4me3 d'un coté de la boucle et de H3K9me, H3K20me et H3K27me3 de l'autre côté. Elle pourrait correspondre à une barrière capable de séparer les domaines actifs de la chromatine des domaines réprimés.

Cross-link de
l'ADN

Coupure avec une
enzyme de
restriction

Marquage des
extrémités à la
biotine

ligation

Purification de
l'ADN avec la
biotine

séquençage

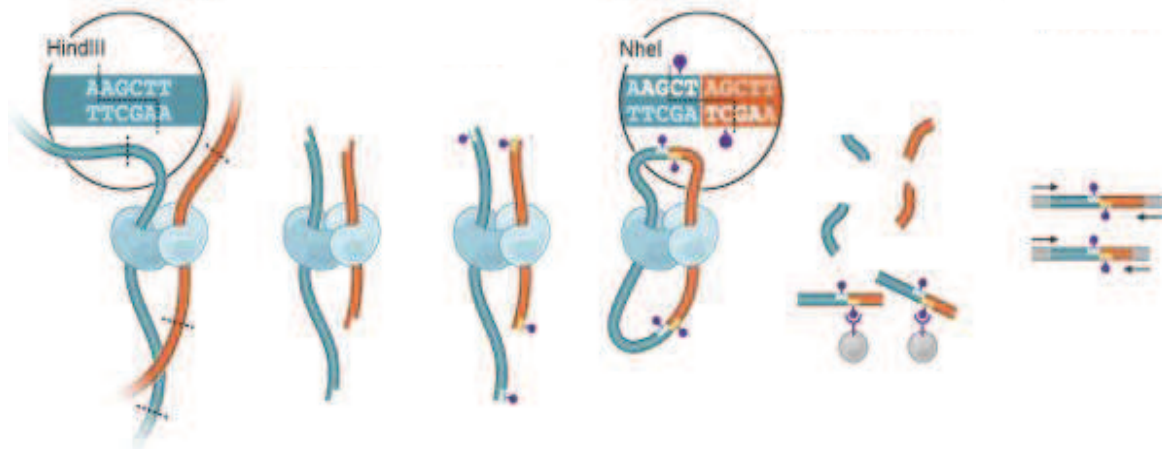


Figure 20: Technique du Hi-C (van Berkum et al., 2010).

Cette technique est un dérivé du 3C analysé ensuite par un séquençage haut débit, permettant d'identifier des interactions à l'échelle du génome sans sélection initiale de régions d'intérêts.

- catégorie V: elle correspond à 27% des boucles de chromatine identifiées mais ne présente pas de statut spécifique d'un point de vue marques de la chromatine.

Un des inconvénients de la technique du ChIA PET est qu'elle se base sur une première étape de ChIP et qu'elle se focalise donc sur la seule sous-population d'interactions entre régions chromatiniques recrutant la protéine d'intérêt sur au moins un côté de la boucle. Du fait de cet enrichissement préliminaire par ChIP, l'ensemble des interactions détectées peut également être biaisé vers des interactions entre fragments génomiques recrutant la protéine d'intérêt des deux côtés de la boucle. Une nouvelle technique mise au point dans le laboratoire de J.Dekker et nommée Hi-C permet quant à elle une identification non biaisée des interactions chromatiniques du génome entier. Cette technique est basée sur celle du 3C, suivi d'un séquençage haut débit des produits obtenus (**Figure 20**). La première étude l'ayant utilisé a identifié de nombreux contacts entre des chromosomes distincts (Lieberman-Aiden *et al.*, 2009) mais a surtout confirmé des données de FISH (Boyle *et al.*, 2001) mettant en avant une interaction préférentielle entre les petits chromosomes riches en gènes, à savoir les chromosomes 16, 17, 19, 20, 21 et 22. De plus, cette étude a permis de définir une division en deux compartiments (A et B) des chromosomes avec des contacts uniquement au sein d'un même groupe et une absence de contact entre groupes distincts. De manière plus générale, le génome lui-même serait divisé en ces deux compartiments, les régions les plus proches sur le génome faisant partie du même groupe. Cette découverte incite à penser que les chromosomes seraient organisés en globules fractaux (Bancaud *et al.*, 2009) qui faciliteraient le déploiement et le repliement de l'ADN lors des différentes étapes cellulaires (activation ou répression de la transcription, cycle cellulaire ...). Une analyse de Hi-C plus récente (Lan *et al.*, 2012) couplée à du ChIP-seq a également montré que les régions impliquées dans la création de boucles de chromatine sont préférentiellement situées à proximité ou incluses au sein de régions de chromatine décondensée. Ceci a permis d'émettre l'hypothèse que les TFs (facteurs de transcription) liés dans ces régions pourraient être à l'origine de l'interactome d'un type cellulaire donné.

III-La transcription

A-Généralités

La transcription est un processus biologique qui permet à partir de l'ADN la création d'une molécule d'ARN et cela, entre autre, dans le but d'aboutir à une protéine à travers un autre mécanisme appelé la traduction. La transcription est un phénomène très finement régulé afin de permettre à une cellule de répondre de façon appropriée à des stimuli externes ou bien internes comme la variation locale en concentration d'une hormone. Il existe différentes étapes dans la transcription, l'initiation, l'élongation et la terminaison qui ont longtemps été considérées comme indépendantes les unes des autres et mettant donc en jeu des facteurs protéiques différents selon les stades. Dans les cellules eucaryotes, 3 ARN polymérases différentes (Burley et Roeder, 1996), l'ARN-PI, l'ARN-PII et enfin l'ARN-PIII réalisent la transcription ce qui permet de distinguer 3 classes de gènes en fonction de la polymérase qui les transcrit :

- Les gènes de la classe I : ils sont à l'origine de la production des ARNs ribosomiques (ARNr) 18S, 28S et 5,8S. Ils sont organisés en répétitions de tandem.
- Les gènes de classe II : ils aboutissent à la formation de protéines *via* leur copie en des molécules d'ARNs messagers, les ARNm. Leur transcription sera le sujet central de cette partie.
- Les gènes de classes III : ils sont la source de production de l'ARN 5S, des ARNs de transfert (ARNt), impliqués dans la traduction de l'ARNm en protéine mais également de divers petits ARN non codant (les ARNnc). Ils sont principalement transcrits pendant l'interphase.

Selon que l'on soit chez les procaryotes ou les eucaryotes, les mécanismes impliqués dans la transcription sont plus ou moins complexes et d'une manière générale sont différents. Nous ne traiterons que du cas des eucaryotes dans ce travail et principalement de la transcription des gènes de classe II qui met en action l'ARN-PII.

B-L'accessibilité de l'ADN

Au sein du noyau, la structure même de la chromatine est peu propice à la liaison de facteurs impliqués dans la régulation de la transcription au niveau de leurs séquences d'ADN

Tableau 2: Sous-unités du complexe de remodelage de la chromatine SWI/SNF dans différentes espèces eucaryotes modèles (Lee et Young, 2000).

Complexes	Organisme	Sous-unité	Complexes	Organisme	Sous-unité
Swi/Snf	<i>S.cerevisiae</i>	Swi1	dSwi/Snf	<i>D.melanogaster</i>	Brahma
		Swi2/Snf2			BAP155
		Swi3			BAP111
		Snf5/Swi10			BAP74
		Snf6			BAP60
		Snf6			BAP55
		Snf11			BAP47
		Snf12			BAP45
		Swp82	hSwi/Snf	<i>H.sapiens</i>	BAF 250
		Swp59			Brg1 ou hBRM
		Swp59			BAF170
		Swp61			BAF155
		Swp29			BAF110
					BAF60a ou 60b ou 60c
					BAF57
					BAF53
	BAF47				
RSC	<i>S.cerevisiae</i>	Sth1			
		RSC1			
		RSC2			
		RSC3			
		RSC4			
	RSC15				
	Sfh1				

cibles. Le remodelage de cette structure peut être assuré par deux types de complexes multiprotéiques. L'activité des complexes du premier groupe utilise l'énergie issue de l'hydrolyse de l'ATP afin d'altérer la position des nucléosomes ou bien de les déstabiliser (Kingston *et al.*, 1996). Ce premier groupe peut être divisé en 5 familles : SWI/SNF ; ISWI ; Mi-2/NuRD ; INO80 et SWR1. Le second groupe de complexes permettant un remodelage de la structure chromatinienne inclut des enzymes capables de modifier directement les histones comme les HATs, HMTs, HDACs ...

1-Les complexes de remodelage ATP-dépendants

a-La famille SWI/SNF

SWI/SNF est la première famille à avoir été identifiée chez la levure *S.cerevisiae*. Ce complexe est composé de 11 sous-unités dont Snf2, Snf5 et Swi1 (Cairns *et al.*, 1994 ; Coté *et al.*, 1994 ; Peterson *et al.*, 1994) (**Tableau 2**). Il est requis pour permettre l'expression de certains gènes comme HO, SUC2 ou INO1 (Biggar et Crabtree, 1999) à travers sa capacité à remodeler la structure chromatinienne durant l'activation de la transcription. Ceci s'effectue par l'intermédiaire de sa sous unité Swi2, seule protéine à présenter une activité ATPasique dans ce complexe (Laurent *et al.*, 1993 ; Coté *et al.*, 1994). D'autres travaux ont également montré que le rôle de Swi/Snf dépasserait le cadre strict de l'initiation de la transcription et serait requis pour les autres étapes de ce processus (Biggar et Crabtree, 1999).

Chez l'Homme, les complexes SWI/SNF présentent une composition relativement hétérogène, certaines de leurs sous-unités différant selon le contexte cellulaire (type cellulaire, état de différenciation, et stade de développement). L'activité ATPasique de ces complexes est portée par deux protéines chez l'Homme : Brg1 (Brm-related gene 1 protein) et hBrm (human Brm) (Khavari *et al.*, 1993). Ces deux protéines, ayant une forte homologie avec Swi2/Snf2, sont retrouvées dans des complexes appelés hSWI/SNFA et hSWI/SNFB (Kwon *et al.*, 1994). Elles ont la capacité à augmenter la fonctionnalité des régions chromatiniennes activant la transcription (Muchardt et Yanic, 1993) en altérant l'organisation nucléosomale de ces séquences (Imbalzano *et al.*, 1994 ; Kwon *et al.*, 1994). Cette modification de la chromatine permet par exemple un meilleur recrutement de TFIIA et de la TBP (TATA binding protein) au niveau de la boîte TATA des promoteurs de certains gènes, une étape déterminante dans l'activation de la transcription (Wilson *et al.*, 1996). C'est l'inclusion différentielle au sein des complexes de type SWI/SNF de certaines protéines telles que les BAFs (Brg1 Associated Factors) possédant des domaines de liaison à l'ADN variables qui

Tableau 3: Sous-unités du complexe de remodelage de la chromatine ISWI dans différentes espèces eucaryotes modèles (Lee et Young, 2000).

Complexes	Organisme	Sous-unité	Complexes	Organisme	Sous-unité
ISW1	<i>S.cerevisiae</i>	Iswi1	ACF	<i>D.melanogaster</i>	ISWI
		p110			Acf1
		p105	CHRAC	<i>D.melanogaster</i>	p175
		p74			P160
ISW2	<i>S.cerevisiae</i>	p140	hACF	<i>H.sapiens</i>	ISWI
		Isw2			p20
NURF	<i>D.melanogaster</i>	NURF-215			p18
		ISWI			hSnf2h
		NURF-55			BAZIA
		NURF-38	RSF	<i>H.sapiens</i>	p325
					hSnf2h

assure leur ciblage sur différents domaines chromatinien (Wu *et al.*, 2009) (**Tableau 2**), même si ce recrutement ne se ferait pas de manière séquence-spécifique mais plutôt *via* la reconnaissance de structures précises (Quinn *et al.*, 1994). Certaines des sous-unités de SWI/SNF possèdent des domaines permettant la reconnaissance spécifique de modifications post-traductionnelles des histones comme l'acétylation des lysines par des bromodomaines (Singh *et al.*, 2007).

b-La famille ISWI

Comme dans le cas des complexes de type SWI/SNF, c'est la sous-unité portant l'activité ATPase, ISWI, qui définit cette famille de complexes perturbant la structure chromatinienne en présence d'ATP (Elfring *et al.*, 1994) (**Tableau 3**). Ainsi, chez la *Drosophile*, trois complexes de ce type ont été identifiés : NURF (Tsukiyama et Wu, 1995), CHRAC (Varga-Weisz *et al.*, 1997) et ACF (Ito *et al.*, 1997) (**Tableau 3**). En plus de sa partie ATPasique, ISWI possède également un domaine de liaison aux histones (principalement à l'histone H4) : le domaine SANT (Swi3, Ada2, N-CoR, TFIIB) qui agit également comme une plateforme d'interaction pour certaines protéines (Boyer *et al.*, 2004) et un domaine SLIDE (SANT Like ISWI Domaine) qui permet des interactions avec l'ADN nucléosomal (Grune *et al.*, 2003). Des homologues de ISWI ont été identifiés chez tous les eucaryotes supérieurs, suggérant un rôle important de cette protéine. Chez l'Homme, deux sous-unité ATPasique, différentes dans leurs extrémités C-terminale et N-terminale, ont été identifiées, il s'agit de SNF2-L et SNF2-H (Okabe *et al.*, 1992 ; Aihara *et al.*, 1998). Ces deux protéines ne sont pas exprimées dans les mêmes tissus, suggérant ainsi des fonctions tissu-spécifiques (Lange *et al.*, 2011), et ne sont pas incluses dans les mêmes complexes de type ISWI. En plus de ces cœurs catalytiques, les complexes ISWI comprennent plusieurs sous-unités alternatives qui non seulement permettent la reconnaissance différentielle de domaines chromatinien particuliers, mais assurent également des fonctions multiples, comme l'activation de la transcription (Strohner *et al.*, 2001), la régulation de l'élongation et de la terminaison de la transcription (Morillon *et al.*, 2003) mais également la régulation de la réplication (de la Serna et Imbalzano, 2002).

C-La famille Mi2-NuRD

Ce complexe de remodelage est, comme pour les deux familles précédentes, particulièrement hétérogène dans sa composition. Il est caractérisé par la présence de la sous-unité Mi-2/CHD (Chromodomain helicase DNA binding protein) à activité ATPase. Elle

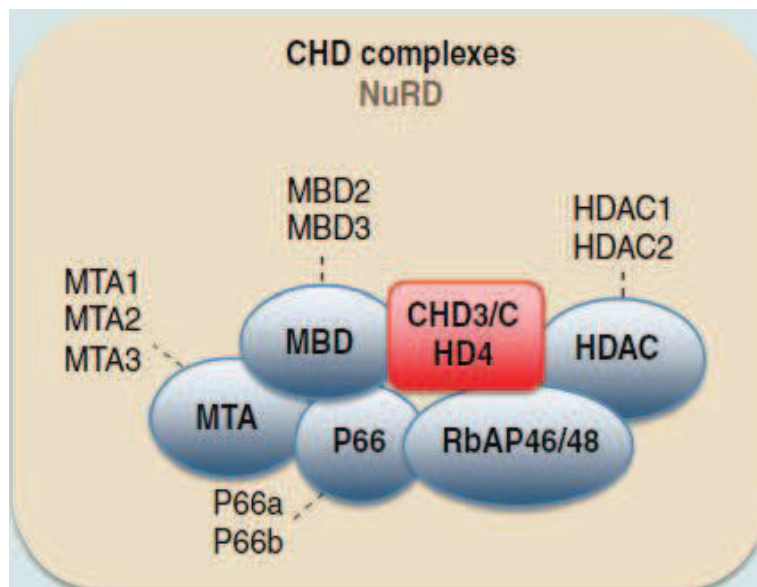


Figure 21: Sous-unités du complexe Mi2/NuRD chez l'Homme (Lange et al., 2011).

Dans ce complexe, la sous-unité CHD, en rouge, porte la fonction ATPasique. Elle peut être CHD3 ou CHD4. Les autres protéines s'assemblent de façon spécifique afin de spécialiser les fonctions de chaque type de complexe.

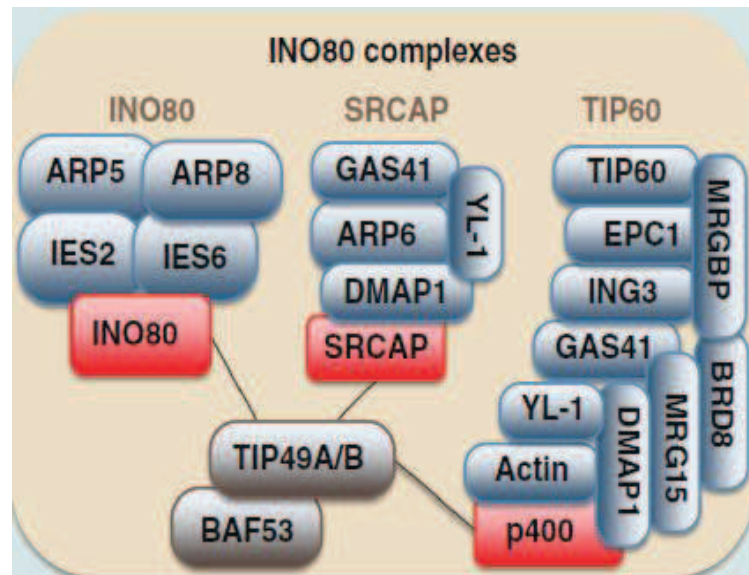


Figure 22: Sous-unités des complexes INO80 chez l'Homme (Lange et al., 2011).

Ce complexe, initialement identifié chez la levure *S.cerevisiae*, est principalement impliqué dans le remodelage de la chromatine au niveau de gènes activés suite à des dommages à la chromatine.

existe sous deux isoformes chez l'Homme : Mi-2 α /CHD3 et Mi-2 β /CHD4 (Woodage *et al.*, 1997) (**Figure 21**). En plus de chromodomaines qui reconnaissent les lysines méthylés (Lange *et al.*, 2011), les sous unités Mi-2 présentent un domaine KRAB (krueppel associated box) impliqué dans la répression transcriptionnelle et dont les fonctions nécessitent le recrutement de la sous-unité KAP-1 (Schultz *et al.*, 2001). Le complexe Mi-2/NuRD comprend également des sous-unités de type MBD, capables de liaison à l'ADN méthylé, les HDAC1 et HDAC2 qui permettent une déacétylation des histones (Humphrey *et al.*, 2001) et enfin des protéines de la famille MTA, MTA1, MTA2 ou MTA3, impliquées chacune dans une réponse transcriptionnelle spécifique (Bowen *et al.*, 2004). L'assemblage spécifique de certaines sous-unités au sein des complexes Mi2-NuRD permet, comme dans le cas des complexes précédents, une spécialisation de leurs fonctions.

d-INO80

La protéine Ino80 a été découverte par homologie de séquence avec l'ATPase ISWI du complexe NURF (Ebbert *et al.*, 1998). Elle appartient à la famille de protéine SNF2/SWI2 et présente un domaine ATPasique très conservé (Tsukiyama *et al.*, 1999) que l'on retrouve chez la Drosophile (dIno80) et chez l'Homme (hIno80). Ino80 fait partie d'un complexe de remodelage de la chromatine, identifié chez *S.cerevisiae*. Ce complexe de remodelage, INO80, possède également une sous unité d'actine G, des protéines Arp4, Arp5 et Arp 8 ainsi que de Rvb1 et Rvb2 (**Figure 22**). Il semblerait que ce complexe soit principalement impliqué dans la réparation de l'ADN à travers le remodelage de la chromatine au niveau de gènes activés par les dommages à l'ADN (Shen *et al.*, 2000).

e-SWR1

Swr1, tout comme Ino80, fait partie de la famille de protéine SNF2/SWI2. Cette protéine est présente en complexe avec une douzaine d'autres protéines dont l'actine G, Arp4, Rvb1 et Rvb2, communes à d'autres complexes de remodelage de la chromatine. On trouve également dans le complexe SWR1 des histones et en particulier l'histone H2A.Z qui établit une interaction à la fois physique et fonctionnelle avec Swr1. Le rôle de ce complexe serait principalement d'assurer l'échange de H2A par son variant H2A.Z, et cela de façon ATP dépendante (Mizuguchi *et al.*, 2004).

2-Les enzymes de modifications des histones et les variant d'histones

Les modifications post-traductionnelles présentent sur les extrémités N-terminales des histones comme l'acétylation ou bien la méthylation de certains résidus sont impliquées dans l'accessibilité de la chromatine aux protéines de la machinerie de transcription. Il en est de même pour l'inclusion de certains variants d'histone comme H2A.Z qui créent des zones chromatinienne incluant des nucléosomes instables, sur lesquelles des facteurs impliqués dans la transcription peuvent être mobilisés. Les protéines impliquées dans les modifications des histones ont été développées dans le paragraphe II-B-1-, et les histones et les variants d'histones ayant un rôle dans la régulation transcriptionnelle ont été décrits dans le paragraphe II-A-2 et II-A-3 respectivement.

C-Activation de la transcription des gènes ARN-Pol II dépendants

1-Généralités: promoteurs et séquences régulatrices.

Il est dans un premier temps nécessaire de pouvoir définir au mieux et selon nos connaissances actuelles les séquences régulatrices de la transcription d'un gène de classe II. Ces gènes, incluant des séquences exons correspondant à des séquences codantes et des introns non codants qui seront éliminés au cours des étapes d'épissage, sont tout d'abord placés sous le contrôle d'un promoteur dit basal que l'on peut restreindre aux séquences les plus proches du TSS. Les plus simples de ces promoteurs ne contiendraient qu'une séquence de type TATA box, riche en A/T et dont le consensus est TATAA (Sandelin *et al.*, 2007) localisée de -25 à -30pb en amont du TSS et/ou un élément initiateur riche en pyrimidine localisé sur le TSS appelé Inr dont la séquence consensus est YYANWYY (Y: C ou T ; N: A/T/C ou G et W: A ou T) avec le A correspondant au +1 du TSS.

On peut ensuite discriminer de ce promoteur basal un promoteur appelé proximal, prenant en compte des séquences d'ADN situées de façon plus lointaine/distale du TSS (entre -50 et -200 pb) avec par exemple la CAAT box (vers -70) et la GC box (entre -80 et -160pb). Ce promoteur peut inclure également des séquences de fixation spécifiques pour des facteurs de transcription régulateurs. Si la distinction entre ces deux types de promoteurs peut être établie en fonction de leur éloignement relatif du TSS, elle présenterait surtout un sens vis à vis de l'initiation de la transcription en elle-même. En effet, la transcription basale serait principalement assurée par des mécanismes initiés au niveau du promoteur basal, sur lequel la

liaison de la TBP et des TAFS (TBP associated factors) permet la mobilisation des facteurs généraux de la transcription (GTFs) au nombre de sept chez les mammifères (Conaway et Conaway, 1993) et de l'ARN-PII pour aboutir à la formation d'un complexe de pré-initiation (CPI). La présence d'éléments régulateurs au niveau du promoteur proximal permettrait quant à elle de moduler cette transcription basale. D'autres séquences régulatrices de la transcription peuvent ensuite être définies, appelées enhancers et silencers dont les rôles respectifs sont de recruter des protéines soit activant ou inhibant la transcription de gènes situés à distance de ces régions. Ces éléments peuvent se retrouver dans des régions intergéniques ou au sein d'introns/exons des gènes qu'ils régulent voire d'autres gènes, ou même situés sur d'autres chromosomes. On peut noter que la définition de ce type d'élément sur des critères de distance en terme de pb rend parfois difficile la distinction sémantique entre promoteur proximal et enhancer/silencer lorsque ces derniers sont situés par exemple dans une fenêtre de 5 kb autour du TSS. D'autres régions régulatrices, appelées insulateurs, permettraient quant à elle d'isoler des séquences données de l'activité transcriptionnelle environnante.

2-Mise en place du complexe de pré-initiation de la transcription

Malgré les progrès effectués *in vivo* ces dernières années, la pleine compréhension des interactions protéines-protéines et protéines-ADN impliquées dans l'initiation de la transcription et ses mécanismes n'est pas encore atteinte. Même si plusieurs cas et modèles d'assemblage du CPI ont été décrits (comme détaillé au sein des paragraphes suivants), quelque soit le modèle, l'initiation de la transcription en elle-même nécessite la présence de nucléotides, une séparation des deux brins d'ADN au niveau du TSS, et la phosphorylation directe ou indirecte du domaine C-terminal (CTD) de l'ARN-PII sur la sérine 5 (Drapkin et Reinberg, 1994).

a-Cas des promoteurs TATA⁺

Deux modèles de recrutement des différents facteurs composant le CPI se font face, un modèle séquentiel ("multistep") et un modèle de type holoenzyme.

Le modèle séquentiel : L'ARN-PII en elle-même n'est pas suffisante pour amorcer le processus de transcription qui nécessite en fait la formation du CPI sur le promoteur (Burley et Roeder, 1996). Dans ce modèle, ceci serait opéré par un assemblage séquentiel de GTFs (Maldonado et Reinberg, 1995). La première étape de l'initiation est le recrutement sur la TATA box de la TBP (Nikolov et Burley, 1991) qui est un facteur initiateur universel car elle

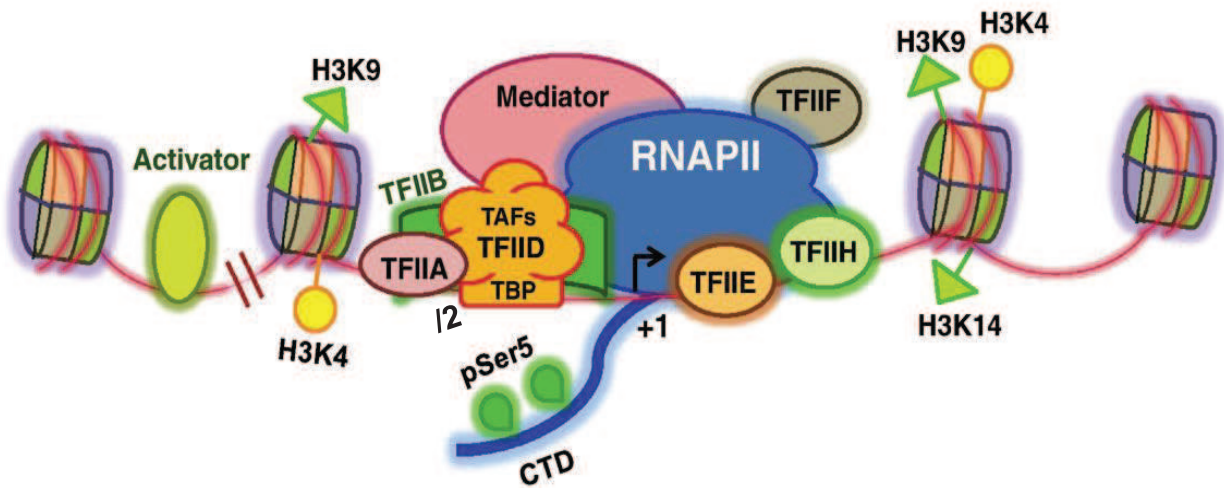


Figure 23; Complexe de pré-initiation (CPI) sur un promoteur TATA+ (Shandilya et Roberts, 2012).

Pour initier le processus de transcription sur un gène ayant une boîte TATA, l'ARN-Pol n'est pas suffisante. D'autres facteurs protéiques tels que la TBP (TATA Binding Protein), les TAFs (TBP Associated Factors) ou encore des GTFs (General Transcription Factors) sont alors recrutés. Au sein de ce complexe, le médiateur permet de faire le lien entre les éléments régulateurs distants et le promoteur basal où sont situés les GTFs et de maintenir sur le promoteur certains cofacteurs activateurs afin de promouvoir d'autres cycles de transcription sur un même gène. La transcription est initiée quand le CTD (C Terminal Domain) de l'ARN-Pol est phosphorylé sur la sérine 2 et la sérine 5.

est impliquée dans les processus d'initiation de la transcription quelque soit l'ARN-P (Hernandez, 1993). La mobilisation des TAFs par la TBP forme le complexe multi protéique TFIID (Cler *et al.*, 2009) qui a la capacité d'engager des interactions physiques et fonctionnelles avec les GTFs. Parmi ceux-ci, le premier à être recruté dans l'ébauche du CPI est TFIIB qui permet un ciblage correct de l'ARN-P sur le promoteur. En effet, sa déplétion dans le système levure entraine un défaut de placement de l'ARN-P sur le TSS (Burley et Roeder, 1996). Le complexe TFIIB-TFIID présent sur le promoteur est ensuite reconnu par un complexe formé de l'ARN-P et de TFIIF, rejoint par TFIIIE, TFIIH et TFIIF mais également par TFIIA dont le rôle serait triple : (i) stabiliser la liaison entre l'ADN et TFIID, (ii) augmenter l'activation de la transcription *via* des activateurs comme Sp1, VP16, NTF1 et Zta (Ozer *et al.*, 1994 ; Yokomori *et al.*, 1994) et enfin (iii) prévenir l'action de facteurs inhibiteurs de la transcription ayant pour cible TFIIB comme NC1 et NC2 (Auble *et al.*, 1994). La transcription est initiée quand la sous unité cdk7 de TFIIH phosphoryle le CTD de l'ARN-P sur les sérines 2 et 5 (**Figure 23**).

Le modèle holoenzyme : ce modèle initialement identifié chez la levure met en avant la présence d'un complexe préexistant dans les cellules et formé de la RNA-P, de TFIIB, TFIIH et TFIIF ainsi que d'autres polypeptides non identifiés (Koleske et Young, 1994). Ce complexe serait recruté directement sur le promoteur après liaison de TBP/TFIID sur la TATA box.

b-Cas des promoteurs TATA⁻

Les promoteurs avec une boîte TATA sont une minorité dans les génomes mammifères: seuls 10 à 20% d'entre eux présentent en effet une TATA fonctionnelle (Gershenzon et Ioshikhes, 2005). Il semblerait que dans ce cas la présence de l'Inr soit suffisante pour diriger la transcription (Weis et Reinberg, 1997 ; Smales *et al.*, 1998). Les résidus -1 et +1 présents dans ces Inr présentent une conservation importante entre les séquences de gènes orthologues ; on y observe un consensus pyrimidine/purine (py/pu) avec principalement des dinucléotides CG, TG et CA (Carninci *et al.*, 2006). Ce consensus py/pu est directement reconnu par TFIID, ce qui permet un recrutement des autres GTFs et de l'ARN-P sur le TSS. De plus pour ces promoteurs sans TATA box, des études ont montré que TFIID pouvait reconnaître le promoteur par l'intermédiaire de ses TAFs. Certains de ces derniers seraient en effet mobilisés sur la chromatine du fait de la présence d'autres séquences promotrices spécifiques (DPE...). De plus, certains des TAFs possèdent des domaines de

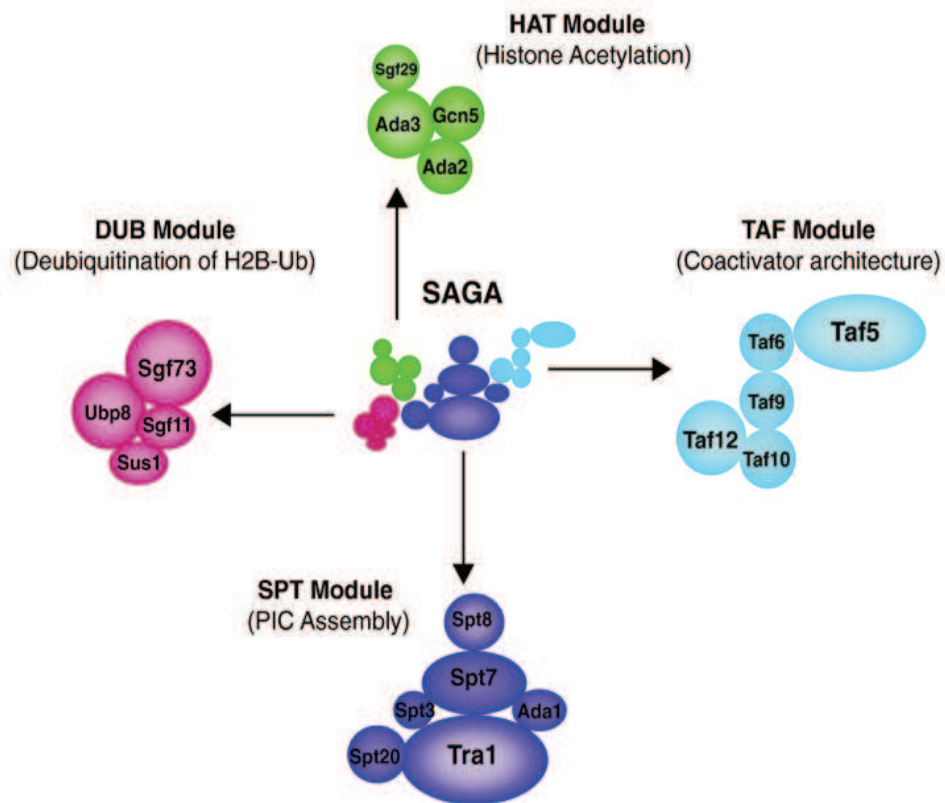


Figure 24: Composition du complexe SAGA (Samara et Wolberger, 2011).

Le complexe SAGA, impliqué entre autre dans l'initiation de la transcription, est composé de 4 domaines ayant chacun plusieurs sous-unités. En vert, le domaine HAT qui permet l'acétylation de plusieurs lysine sur l'histone H3 *via* Gcn5, en bleu clair, le domaine TAF, en bleu foncé, le module SPT et enfin en rose, le domaine DUB qui permet de façon indirecte la phosphorylation du CTD de l'ARN-Pol.

reconnaissance spécifiques pour des modifications post-traductionnelles histones telles que H3K4me3 et H3K9ac, généralement associées à des promoteurs actifs. C'est par exemple le cas des bromodomaine de TAF1 et homéodomaine de TAF3, ou encore le domaine « WD40 repeat » de TAF5. Ces TAFs seraient donc susceptibles de permettre un recrutement indirect et/ou une stabilisation du complexe TFIID sur ces promoteurs TATA.

c-Initiation *via* le complexe SAGA

D'autres études ont mis en avant le rôle du complexe SAGA qui aurait la capacité à remplacer TFIID au niveau de certains promoteurs afin de permettre la formation du CPI. Il serait par exemple impliqué dans l'activation de 10% des gènes chez la levure, principalement liés à la réponse aux stress environnementaux (Baker et Grant, 2007 ; Rodriguez-Navarro, 2009 ; Koutelou *et al.*, 2010). Le complexe SAGA est constitué de 21 protéines, conservées de la levure à l'Homme (Koutelou *et al.*, 2010) et organisé en 4 sous-complexes ayant chacun une fonction bien définie : le module DUB, capable d'enlever un résidu ubiquitine sur la lysine 123 de l'histone H2B (Daniel *et al.*, 2004) *via* Ubp8 permettant ainsi le recrutement de Ctk1 dont le rôle est de phosphoryler le CTD de l'ARN-Pol II (Wyce *et al.*, 2007) ; le module HAT dont le rôle est d'acétyler les résidus H3K9, H3K14, H3K18 et H3K24 (Daniel *et al.*, 2005) *via* Gcn5 ; et enfin les modules SPT et TAF, impliqués dans le recrutement du CPI (Samara et Wolberger, 2011) (**Figure 24**). SAGA a la capacité d'interagir avec la chromatine par l'intermédiaire d'une protéine appelée Sgf73, qui fait partie du module DUB (Bonnet *et al.*, 2010). Cette mobilisation est cependant conditionnée par la méthylation préalable de l'H3K4 qui est reconnue par le chromodomaine d'une protéine de SAGA (Lee *et al.*, 2011). Cette interaction pourrait également jouer un rôle primordial dans la régulation de la fonction HAT de Gcn5. Il semblerait que SAGA soit également impliqué dans l'élongation de la transcription (Govind *et al.*, 2007) ainsi que dans l'export des ARNm nouvellement synthétisés par l'intermédiaire de la protéine Sus 1, appartenant également aux pores nucléaires (Pascual-Garcia *et al.*, 2008).

3-Modulation de l'initiation par les promoteurs proximaux:

a-Assemblages différentiels

L'assemblage différentiel du CPI en fonction du promoteur est un paramètre important de régulation, qui pourrait premièrement influencer sur la dynamique structurale de ce complexe qui peut présenter une conformation ouverte ou fermée de sa cavité centrale dont les rôles

respectifs dans la fonction du CPI ne sont pas encore élucidés (Cler *et al.*, 2009). Cet assemblage différentiel peut être en partie relié aux éléments de régulation portés par les promoteurs proximaux, certains TAFs étant impliqués préférentiellement dans la médiation de l'activité régulatrice de facteurs de transcription spécifique. C'est par exemple le cas du facteur NT1 dont l'action nécessite en plus de la TBP, TAFII250 et TAFII150 alors que celle Sp1 requiert en plus de ces 3 protéines TAFII110 (Chen *et al.*, 1994). De plus, dans le cas de promoteurs TATA⁻, Sp1 serait également capable d'interagir avec TAF4 pour recruter TFIID (Gill *et al.*, 1994).

b-Le complexe Mediator

Lors d'évènements de régulation transcriptionnelle médiés par le recrutement de facteurs de transcription (TFs) au niveau du promoteur proximal ou d'éléments distants de type enhancers, les informations apportées par ces protéines doivent être transmises aux GTFs présents sur le promoteur basal afin d'en modifier la fonction. C'est un des rôles du complexe multi-protéique Mediator caractérisé dans un premier temps comme complexe holoenzyme alternatif chez la levure incluant l'ARN-Pol II et les protéines SRB, Gal II, Sug 1 et TFIIF (Kim *et al.*, 1994). Des travaux ultérieurs ont permis de mieux caractériser ce complexe, composé de 21 sous-unités, divisible en trois sous-modules (Davis *et al.*, 2002) : la tête et le corps, impliqué dans l'interaction avec la RNA-Pol II et la queue, plateforme pour des interactions avec des TFs. L'établissement de ce pont peut être facilité par une modulation locale de la conformation de l'ADN due à la liaison de la TBP au niveau du promoteur basal. En effet, cette dernière se fixe dans le petit sillon de la chromatine, ce qui entraîne une déformation importante pour la coordination et la stabilisation du CPI mais également pour le rapprochement spatial du CPI avec les TFs liés sur les enhancers (Burley et Roeder, 1996). Il faut noter que si Mediator n'a été envisagé qu'en tant que co-activateur de la transcription, du fait de ses interactions avec l'ARN-Pol II et de sa capacité à promouvoir la formation du CPI ainsi que son maintien entre les différents 'round' de transcription (Reeves and Hahn, 2008), d'autres travaux ont indiqué qu'il pouvait également être impliqué dans l'inhibition de gènes. En effet, dans le modèle levure, Mediator peut être co-recruté avec Cdk8 (Liu *et al.*, 2001 ; Samuelson *et al.*, 2003) dont la perte entraîne une dérégulation d'un ensemble de gènes. Cependant, la présence de Cdk8 et l'inhibition de la transcription des gènes ne semblent pas

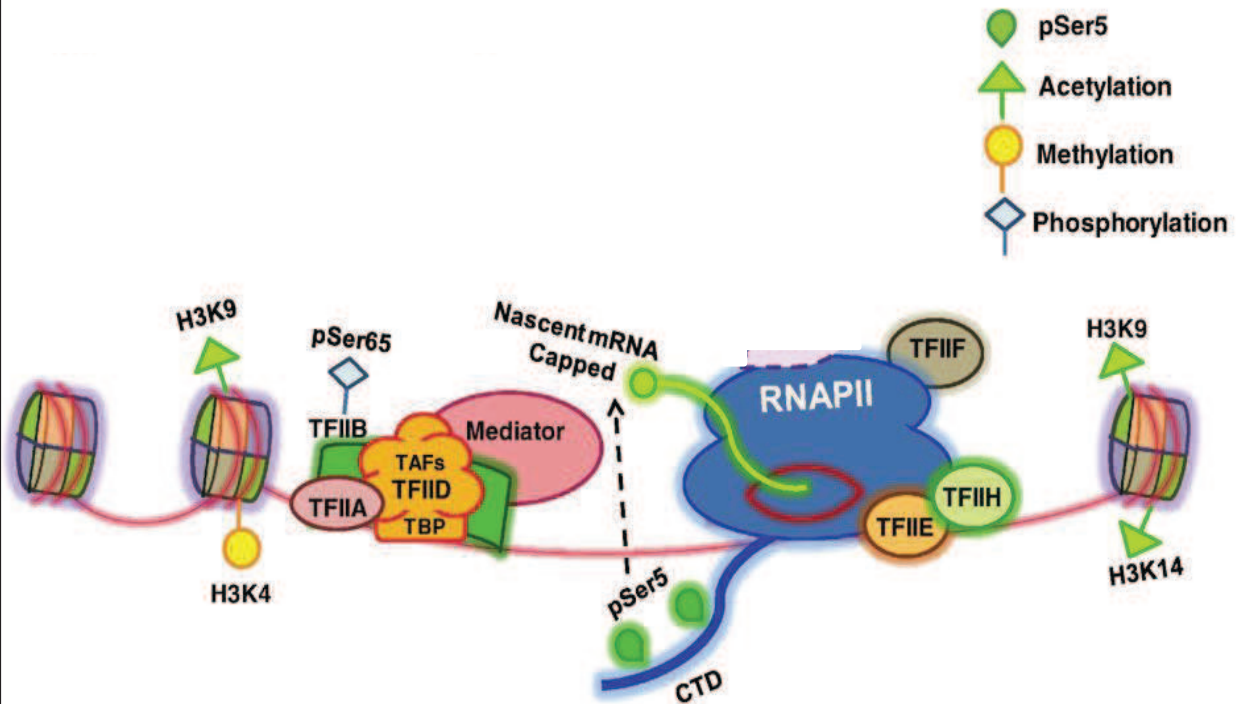


Figure 25: Cycle productif de la transcription (Shandiliya et Roberts, 2012).

La liaison des tous les facteurs nécessaires à la transcription est suivie d'une étape dite de « clearance » où l'ARN-Pol avance et synthétise les premiers nucléotides du futur ARNm. Lorsque 8 nucléotides sont synthétisés, le cycle est dit productif. L'ARN néo synthétisé est alors coiffé pour ne pas être digéré par des nucléases. L'ARN-Pol rentre ensuite dans une phase d'élongation également soumise à des régulations.

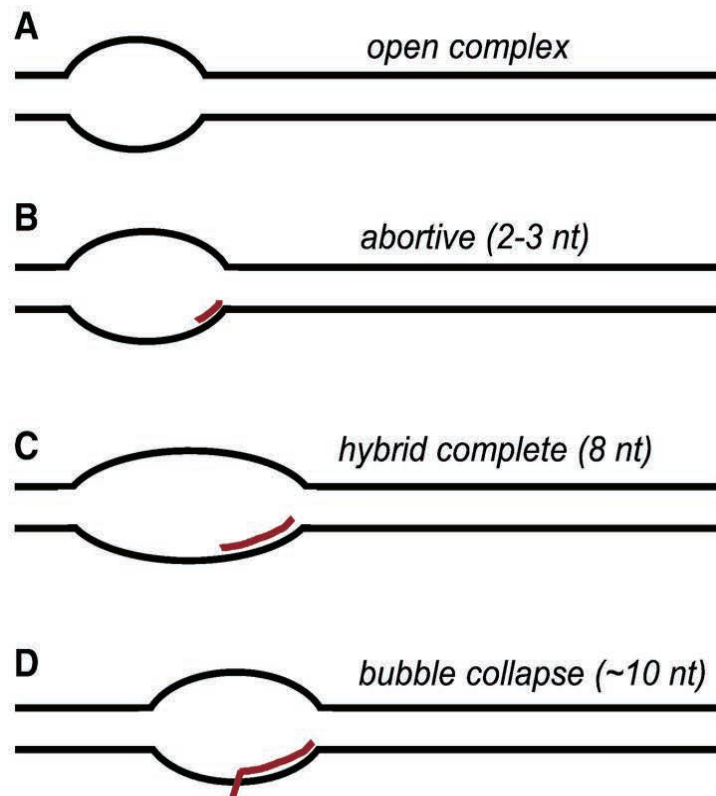


Figure 26: Schéma des différents stades de l'initiation de la transcription (Luse, 2012).

(A) La première étape de la transcription est une ouverture de la double hélice de l'ADN.

(B) Si l'ARNm néo-synthétisé par le complexe de transcription est très court (inférieur à 8 nts), le cycle peut alors être abortif.

(C) Le cycle de transcription sera productif quand l'hybride ADN/ARNm est supérieur à 8-mer.

(D) Le cycle productif entraîne un effondrement de la bulle de transcription en 5' qui marque la fin du phénomène de clearance.

être des événements corrélés chez les mammifères (Wang *et al.*, 2005), ainsi que cela a été suggéré par d'autres études plus récentes chez la levure (Andrau *et al.*, 2006).

4-Elongation et terminaison de la transcription

a-Synthèse de l'ARNm

Le processus d'élongation de la transcription est un phénomène très dynamique soumis à de nombreuses régulations. Il débute par une étape de « clearance » au cours de laquelle le complexe contenant l'ARN-Pol II quitte le promoteur. Ce phénomène est assez bien défini pour des gènes TATA⁺ qui ont été les plus étudiés. La fusion (ou le désappariement) des deux brins nécessaire au début de la synthèse d'un ARNm ne se fait pas au niveau du TSS mais en amont, vers le nt -9. Cela entraîne la formation d'une « bulle » entre cette position qui va rester fixe et le nucléotide en position -2. La formation puis l'avancée de cette bulle en 3' sont des phénomènes centraux de la « clearance ». La synthèse des premiers nucléotides de l'ARNm à cette phase de l'élongation est une étape qui reste critique tant que l'ARNm n'atteint pas une certaine taille. En effet, l'hybride ADN/ARN formé au départ est trop petit pour être stable et peut être relâché du complexe de la transcription. On parle alors de cycle abortif. Il faut noter que l'ARN-Pol II possède cependant une tendance faible à entrer dans ce type de cycles abortifs du fait de la présence à cette étape de TFIIF et de sa sous-unité hélicase XPB qui limite les ré-appariements de l'ADN (Dvir *et al.*, 1996 ; Dvir *et al.*, 1997 ; Moreland *et al.*, 1999 et Lin *et al.*, 2005). La transcription est dite productive lorsqu'un ARN de 8-mer a été synthétisé (**Figure 25**). On observe alors un effondrement de la bulle en 5' qui revient à une taille initiale à savoir une fusion des deux brins sur environ 10 nt. Cet effondrement marque la fin du phénomène de clearance (Luse, 2012) et de la nécessité de la présence de TFIIF pour les fonctions de l'ARN-Pol II (Pal *et al.*, 2005) (**Figure 26**).

Suite à la « clearance », l'élongation de la transcription peut être régulée alors que la polymérase est en cours de synthèse, et notamment lorsqu'elle atteint les 30 à 50 nts en aval du TSS. A ce niveau, dépendant des gènes modèles étudiés, il a été effectivement observé que l'ARN-Pol II pouvait revenir en arrière ou bien s'arrêter (Pal *et al.*, 2001 ; Újvári *et al.*, 2002). Certains complexes protéiques peuvent influencer sur les pauses de l'ARN-Pol II. C'est le cas de DSIF et de NELF qui ont un rôle inhibiteur sur l'élongation (Conaway et Conaway, 2012). Le rôle exact de DSIF est toutefois délicat à définir car, dans certains contextes et principalement en présence de P-TEFb, il semblerait au contraire diminuer le nombre de pauses de l'ARN-Pol II (Conaway et Conaway ; Hartzog et Fu, 2012). Le facteur TFIIS et l'elonguine sont eux au

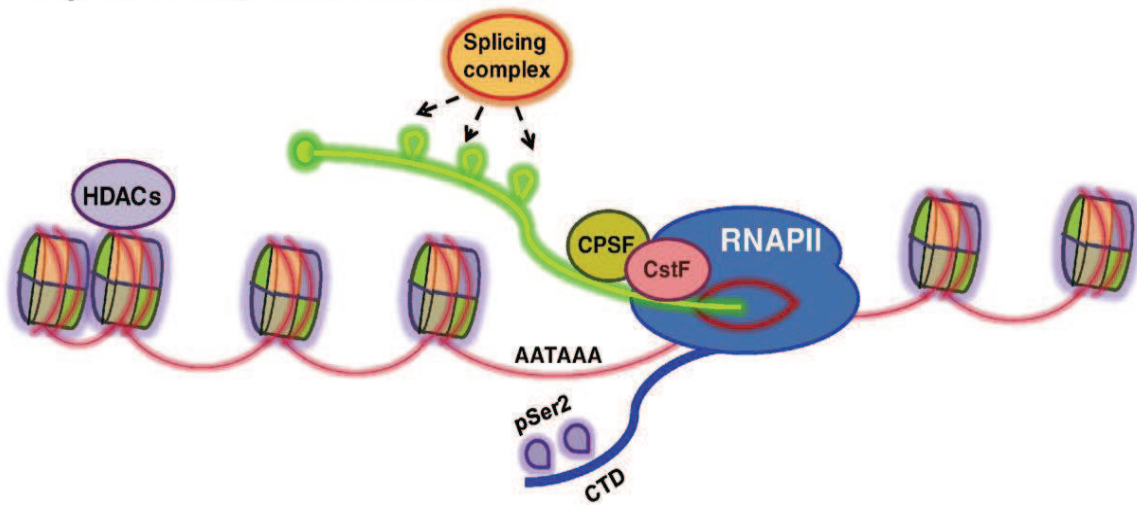


Figure 27: **Schéma illustrant la terminaison de la transcription (Shandilya et Roberts, 2012).**
 La dernière étape de la transcription est la terminaison au cours de laquelle les facteurs CPSF et CstF se lient à la machinerie de transcription afin de cliver l'ARNm néo-synthétisés.

contraire impliqués dans la processivité de l'ARN-Pol II (Fish et Kane, 2002 ; Shandilya et Roberts, 2012). A la suite de cet arrêt, si la machinerie de transcription repart, elle a définitivement acquis toutes les caractéristiques d'un complexe d'élongation mature (Luse, 2012).

De nombreux autres facteurs de transcription sont impliqués dans des phénomènes de régulation de l'élongation qui sont le plus souvent cellule- et gène-spécifiques. La description de cette grande diversité de phénomènes n'étant pas essentielle à la compréhension du sujet sur lequel a porté cette thèse, je ne ferai que souligner l'intervention dans ces régulations de chaperonnes d'histones, qui permettent l'éviction des nucléosomes pendant l'élongation, ainsi que le rôle primordial du CTD de l'ARN-Pol II qui se comporte comme une plateforme de recrutement pour différents facteurs protéiques associés à la régulation de l'initiation, de l'élongation, de la terminaison et bien évidemment de la maturation des ARNs. Elle permet en effet de moduler le recrutement de l'enzyme impliquée dans l'addition de la coiffe à l'extrémité 5' de l'ARN, un phénomène qui s'opère en parallèle de l'élongation (Hahn, 2004).

b-Terminaison

Chez les eucaryotes, la majorité des gènes codant une protéine présente un signal de poly(A) conservé : 5'-AATAAA-3' suivi d'une séquence riche en G/T. La transcription du signal de poly(A) entraîne une diminution de la processivité de l'ARN-Pol II qui se met alors en pause, entraînant un clivage de l'ARN nouvellement synthétisé. Chez l'Homme, le clivage de l'ARN au niveau du site de polyadénylation est dépendant d'un CTD intact (Lee et Young, 2000) requis pour la liaison des facteurs CPSF (cleavage polyadenylation specificity factor) et CstF (cleavage stimulation factor) (**Figure 27**). Une fois la transcription terminée, la déphosphorylation des serines 2 et 5 du CTD de l'ARN-Pol II entraîne son "relargage" du gène. Ceci est catalysé par plusieurs enzymes, Fcp1 au niveau de la serine 2 (Archambault *et al.*, 1998) et de SCP1 pour la ser5 (Meinhart *et al.*, 2005). Des résultats obtenus par différents laboratoires ces dernières années font émerger un modèle de recyclage de l'ARN-Pol II après qu'elle ait transcrit un gène. Ce recyclage ferait intervenir la formation ou la pré-existence de boucles chromatiniennes entre les régions terminales et promotrices des gènes actifs. En effet, TFIIB, présent au niveau du promoteur en tant que GTF, est capable d'interagir avec les facteurs de terminaison CstF et CPSF. Cette boucle permettrait une ré-initiation rapide de la

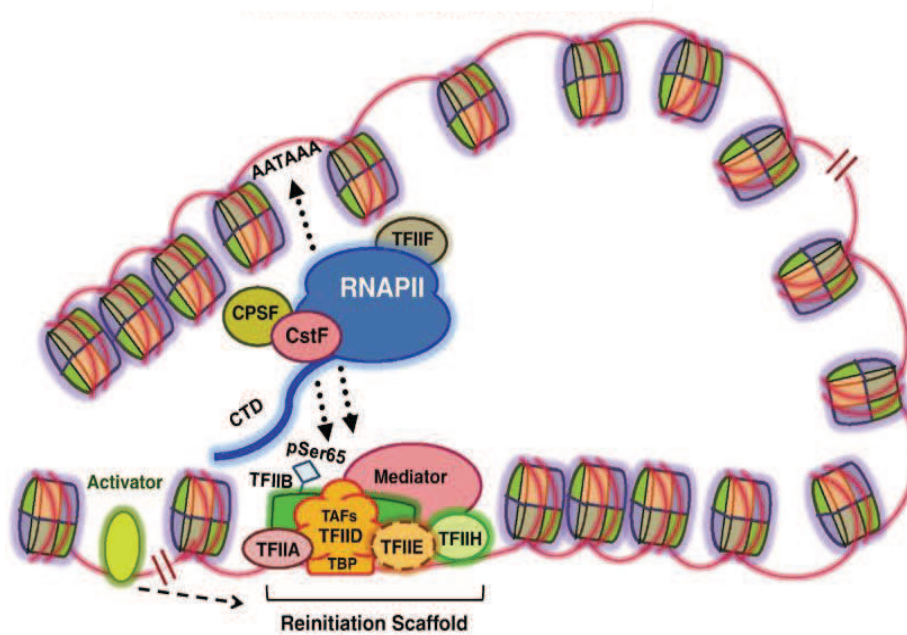


Figure 28: Boucle de chromatine entre le promoteur et le site de terminaison de la transcription d'un gène transcrit (Shandilya et Roberts, 2012).

Plusieurs études menées dans des laboratoires indépendants ont montré l'existence d'une boucle de chromatine entre le promoteur d'un gène et son TTS (Transcriptional Termination Site) et cela principalement dans le but de recycler l'ARN-Pol pour une nouvelle initiation rapide de la transcription. Cette boucle se ferait à travers les GTFs et les facteurs de terminaison CPSF et CstF.

transcription sur les gènes activés (Calvo et Manley, 2003 ; Singh et Hampsey, 2007) (**Figure 28**).

5-La dynamique de la transcription

Depuis les premiers travaux effectués sur la transcription, les connaissances ont évolué ; contrairement à ce qui a été pensé, le recrutement des facteurs impliqués dans la transcription est très dynamique et ceci à deux niveaux : les protéines impliquées dans la transcription et sa régulation (TFs et GTFs) présentent des mobilités intranucléaires très rapides, et les recrutements de ces protéines sur leurs promoteurs sont également des événements dynamiques et même cycliques dans certains cas. La bibliographie concernant les premiers de ces processus étant extrêmement vaste, sur des modèles cellulaires et protéiques différents ainsi qu'à des résolutions spatio-temporelles différentes, j'ai choisi d'illustrer ce point par quelques exemples ayant trait aux deux questions principales sous-jacentes à cette problématique.

a-Dynamique des protéines

La transcription d'un gène peut requérir un recrutement précis des protéines régulatrices impliquées ainsi que des composants de la machinerie de transcription (Metivier *et al.*, 2003 ; Sims *et al.*, 2004 ; Rochette-Egly, 2005 ; Gorski *et al.*, 2008). Cependant, les sites de liaisons spécifiques des TFs étant rares au sein du génome comparé aux nombres de sites de liaisons non spécifiques, se pose donc la question de savoir comment un TF peut trouver et lier rapidement un site spécifique. Les premières études faites sur le génome estimaient que celui-ci ainsi que les protéines y évoluant étaient assez statiques et que les rencontres entre les TFs et leurs sites de liaison se faisaient au hasard par le biais d'une simple diffusion (Gorski *et al.*, 2006). Cette hypothèse n'étant pas entièrement compatible avec des observations faites sur la mobilité intranucléaire des protéines par FRAP (Fluorescence recovery after photobleaching), un autre modèle a été proposé, celui de la diffusion facilitée (Hager *et al.*, 2009 et Normanno *et al.*, 2012). Ce modèle montre une dynamique des TFs dans le noyau présentant deux caractéristiques. La première est une diffusion libre dans l'espace, ne nécessitant pas d'énergie, durant laquelle les TFs n'interagissent pas avec l'ADN. Cette première propriété permet aux TFs d'explorer rapidement l'ensemble du noyau, à une vitesse de l'ordre de 0,5 à 5 $\mu\text{m}^2/\text{s}$, selon les tailles des protéines et leurs capacités à interagir avec l'ADN (Gorski *et al.*, 2006). La seconde caractéristique est une liaison non spécifique des TFs sur l'ADN qui serait suivie d'une exploration de la portion du génome alentour sur une

cinquantaine de bases afin de trouver une séquence cible ("sliding" et "hopping"; Hager *et al.*, 2009). Il est maintenant établi que lors de leur recherche de séquences cibles, les TFs sont particulièrement dynamiques mais, qu'en est-il une fois ce site fixé ?

b-Recrutement sur les promoteurs

La majorité de nos connaissances sur la présence de TFs comme les RNs ou l'ARN-Pol II liés à leurs séquences cibles nous viennent d'expériences telles que le ChIP qui, bien que nous ayant apportés beaucoup de données, ne permettent pas d'appréhender des processus comme la transcription en temps réel. De nouvelles techniques, basées sur la microscopie ont été développées ces quinze dernières années. Elles permettent, entre autre, d'analyser la dynamique des protéines dans des cellules vivantes. Alors que tout le travail effectué *in vitro* ou bien sur des cellules fixées indiquait, une fois le complexe ADN/TF formé, une grande stabilité, de l'ordre de plusieurs heures comme par exemple pour le récepteur aux glucocorticoïdes (GR) qui aurait une demi vie sur l'ADN de 108 min (Perlmann *et al.*, 1990), des études de FRAP faites dix ans plus tard sur des cellules vivantes ont montré que cette liaison est en réalité très dynamique (McNally *et al.*, 2000). Il en est de même pour d'autres TFs comme le récepteur de la progestérone (Rayasam *et al.*, 2005) ou encore aux hormones oestrogéniques (Sharp *et al.*, 2006).

Un des exemples les plus documentés de la dynamique de régulation de la transcription des gènes est celui par les œstrogènes (E2). La vision traditionnelle de la transcription stimulée par E2 était que le récepteur aux œstrogènes ER se liait à son élément de réponse au niveau du promoteur de ses gènes cibles et qu'il restait associé à ce promoteur tant que le signal œstrogénique était présent (Hahn, 1998 ; Berk, 1999). Des travaux initiaux, réalisés par deux laboratoires (Shang *et al.*, 2000 ; Métivier *et al.*, 2003) ont montré que les co-activateurs de la transcription médiée par ER, en présence d'E2, tels que SRC-1, TIF2, GRIP1, AIB1, CBP, p300 et pCAF, ainsi qu'ER lui-même ne sont pas statiques tout le long de la transcription. Au contraire, ces études utilisant majoritairement la technique du ChIP ont montré que ER n'est recruté que 15 minutes suivant le signal E2 et que ce recrutement augmente jusqu'à 30-45 minutes avant de disparaître (ou n'être plus détectable) à partir de 75 minutes ; et que ce cycle se répète avec les mêmes écarts de temps. Ces travaux ont permis de mettre en avant un recrutement cyclique et transitoire des co-activateurs de ER lors de l'activation d'un gène, un processus oscillatoire également décrit pour d'autres TFs (récepteurs aux androgènes, Kang *et*

al., 2002; à la vitamine D, Vaisanen *et al.*, 2005; aux hormones thyroïdiennes, Liu *et al.*, 2006; ou encore aux acides rétinoïques, Bruck *et al.*, 2009). Il faut noter que ces travaux, contrairement aux études se basant sur des techniques d'imagerie sur cellules uniques, caractérisent des événements ayant lieu à l'échelle de populations cellulaires et ne déterminent en aucun cas des temps de résidence pour les protéines suivies. Les comportements oscillatoires observés seraient en fait une résultante des phénomènes rapides de liaison des protéines à l'ADN en "hit and run" et d'une détection de ces événements de liaison ayant lieu au sein d'un nombre plus ou moins important de cellules (Lemaire *et al.*, 2006; Hager *et al.*, 2009; Stavreva *et al.*, 2012).

D-La transcription aux enhancers

Les enhancers sont des éléments présents sur l'ADN qui peuvent activer la transcription à distance en communiquant avec le promoteur (Geyer *et al.*, 1990 ; Sanyal *et al.*, 2012). Ils permettent entre autre le recrutement de certains TFs cellules spécifiques, de l'ARN-Pol II et de facteurs impliqués dans la régulation de l'initiation de la transcription comme Mediator. Le contact entre l'enhancer et un promoteur donné se fait par l'intermédiaire d'une boucle de chromatine (Blackwood et Kadonaga, 1998 ; Bulger et Groudine, 1999 ; de Laat *et al.*, 2008). Ce modèle de fonctionnement a été confirmé par des expériences de 3C (Dekker *et al.*, 2002 ; de Wit et de Laat, 2012) par exemple au niveau des gènes inclus dans le locus de la β -globin où le LCR contenant plusieurs enhancers contacte les gènes de ce locus de façon coordonnée dans le temps (Tolhuis *et al.*, 2002). De manière générale, la formation de ces boucles entre enhancers et promoteurs de gènes régulés semble nécessiter différentes protéines en fonction des cellules et des gènes (Vakoc *et al.*, 2005 ; Jing *et al.*, 2008 ; Kim *et al.*, 2009).

Les enhancers sont présents dans des régions pauvres en nucléosomes et enrichies en variants d'histones tels H2A.Z et H3.3 dont l'incorporation déstabilise la liaison des nucléosomes à l'ADN (Jin et Felsenfeld, 2007 ; Mito *et al.*, 2007 ; Jin *et al.*, 2009) ainsi qu'en certaines marques histones comme H3K4me1 (Bernstein *et al.*, 2002 ; Heintzman *et al.*, 2007 ; Koch *et al.*, 2007), H3K27ac ou H3K27me3 selon qu'ils soient actifs ou inactifs respectivement (Creyghton *et al.*, 2010). Les modifications H3K9me1/me2 et me3 peuvent également être présentes. Les enhancers sont également caractérisés par la présence de certaines protéines

comme p300 ou bien CBP qui par leurs activités HAT permettent le recrutement d'une partie de la machinerie de transcription (Merika *et al.*, 1998).

L'avènement des techniques de RNA-seq et ChIP-seq ont permis d'identifier clairement l'existence de transcription dans des régions intergéniques, à l'origine d'ARNs régulateurs (Birney *et al.*, 2007 ; Prasanth et Spector, 2007). Ces transcrits sont généralement peu abondants, poly-adénylés et non épissés (De Santa *et al.*, 2010), et ont été identifiés au niveau d'éléments *cis* régulateurs comme le LCRs du locus β -globine par exemple où leur rôle est de maintenir une chromatine ouverte (Ashe *et al.*, 1997 ; Gribnau *et al.*, 2000). Le développement de la technique de GRO-seq (global nuclear run on-sequencing) (Core *et al.*, 2008) a permis d'étendre ces observations de transcription extra-génique au niveau des enhancers. En particulier, cette technique a permis d'observer des synthèses rapides (après 10 min de traitement) d'ARNs au niveau des enhancers recrutant le récepteur aux œstrogènes (ER) (Hah *et al.*, 2011). La production de ces eRNA (Kim *et al.*, 2010) apporte un nouveau niveau de régulation de la transcription de manière générale, ces eRNA pouvant jouer des rôles positifs ou négatifs sur la stabilité de l'organisation chromatinienne locale à l'enhancer, la stabilité de l'ARNm produit au niveau de l'autre extrémité des boucles reliant enhancers aux promoteurs de gènes régulés, etc...

E-La répression transcriptionnelle:

1-La répression par les protéines polycomb

Si, comme on l'a vu l'activation des gènes est un phénomène très complexe et finement régulé, il en est de même pour la répression des gènes. Les protéines de type Polycomb (PcG), ont été identifiées comme étant des facteurs répresseurs des gènes homéotiques chez la Drosophile. De fait, dans cette espèce ainsi que chez les vertébrés, les complexes protéiques formés par les PcG permettent la régulation de nombreuses cibles, principalement des gènes impliqués dans le développement et la différenciation cellulaire. Chez la Drosophile, il existe plusieurs complexes PcG: PRC1, PRC2, PhoRC, dRAF et PR-DNB (Beisel et Paro, 2011), et seuls PRC1 et PRC2 sont conservés chez les vertébrés mais ils présentent des variations dans leur composition selon le type cellulaire et le stade de développement. Ces deux complexes sont recrutés au niveau de séquences appelées les PcG responsive elements, PREs qui ne présentent pas de séquences consensus identifiées mais

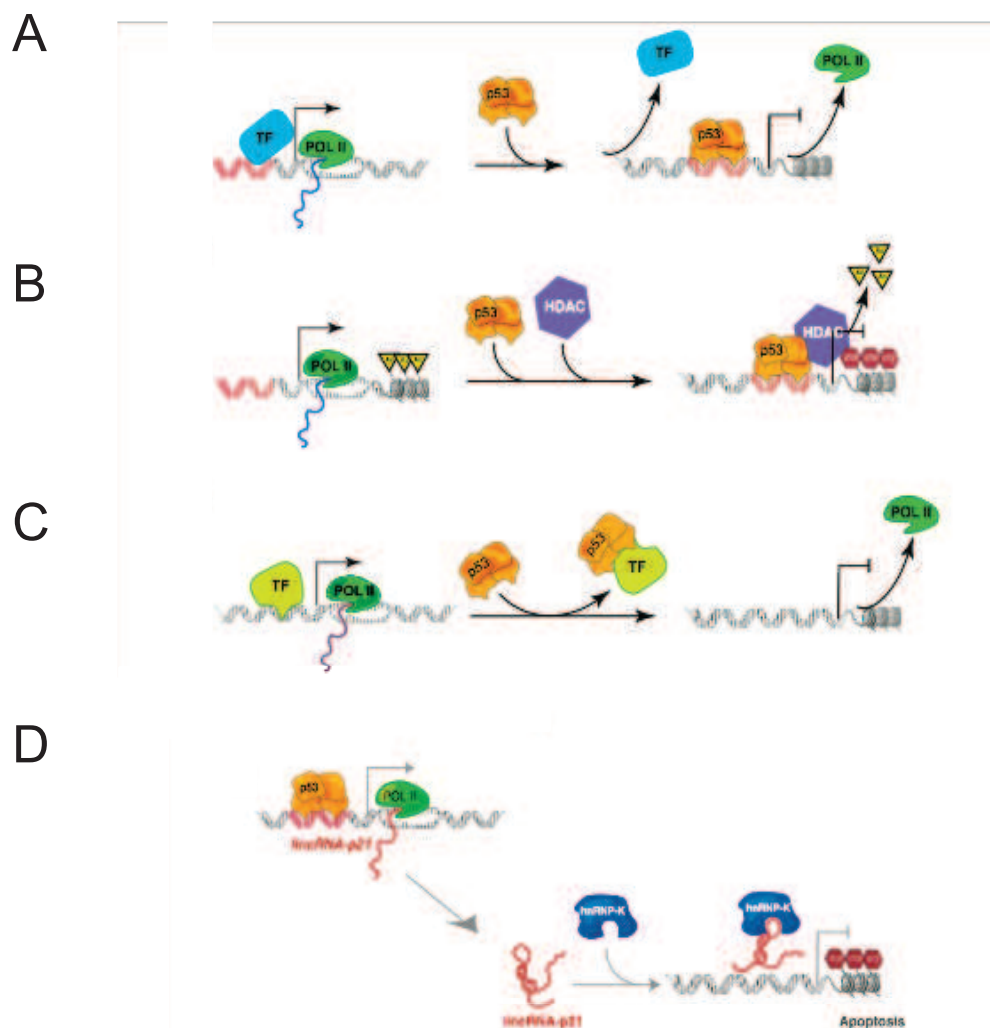


Figure 29: Répression transcriptionnelle par p53 (Rinn et Huarte, 2011).

La protéine p53 peut réprimer la transcription de centaines de gènes, *via* des mécanismes directs ou indirects,.

(A) p53 est capable d'entrer en compétition avec des facteurs activateurs de la transcription

(B) p53 peut également permettre le recrutement d'HDAC ou de complexes répresseurs modifiant la conformation de la chromatine

(C) p53 peut séquestrer des facteurs protéiques comme le récepteur aux glucocorticoïdes (GR).

(D) p53 peut également réguler la transcription de gènes cibles indirects en stimulant la production d'ARNs non codants inhibiteurs comme des micro ARNs (miARNs) ou encore des ARNs intergéniques non codants (lincARNs).

chez les vertébrés, ces séquences sont très corrélées à une forte densité en CpG (Ku *et al.*, 2008 ; Mohn *et al.*, 2008). L'activité répressive de ces complexes s'opère par l'intermédiaire de modifications post-traductionnelles d'histones associées à la chromatine inactive (Bantignies et Cavalli, 2011). Par exemple, le complexe PRC2 *via* sa sous unité E(z) dépose des marques caractéristiques telle que H3K27me3 qui sert de plateforme de recrutement pour PRC1. PRC1 quand à lui est responsable du dépôt d'un groupement ubiquitine sur H2AK119 *via* ses sous-unités RING1B et BMI1 (Cao *et al.*, 2005). La présence de cette modification post-traductionnelle est corrélée à l'inhibition de l'élongation de la transcription par l'ARN-Pol II (Stock *et al.*, 2007 ; Zhou *et al.*, 2008). Enfin, la forte densité en CpG des PREs est associée à une coopération entre systèmes PcG et la méthylation des cytosines qui viendrait compléter/maintenir une répression stable de certains gènes (Ku *et al.*, 2008 ; Mohn *et al.*, 2008).

2-La répression par p53

P53 est une protéine particulièrement importante qui possède un rôle primordial dans le maintien de l'intégrité cellulaire: plus de 50% des cancers chez l'Homme ont une mutation ou bien une délétion sur le gène *p53* (Beckerman et Prives, 2010). Cette protéine a la capacité à réguler de façon coordonnée plusieurs centaines de gènes en se liant sur l'ADN au niveau de séquences qui peuvent être localisées dans le promoteur proximal, dans le 1er intron ou bien plus en aval sur le gène (Menendez *et al.*, 2009). Quinze % des cibles de ce facteur, impliquées dans divers processus importants pour la cellule comme la prolifération, le contrôle du cycle cellulaire, l'apoptose...(Riley *et al.*, 2008), subissent une répression transcriptionnelle, tels les gènes *survivin* (Hoffman *et al.*, 2002 ; Raj *et al.*, 2008) ou *C.Myc* (Ho *et al.*, 2005). L'action répressive de P53 peut rendre compte de plusieurs mécanismes directs ou indirects (Ogbourne et Antalis, 1998). Ainsi, P53 aurait la capacité

d'entrer en compétition avec des facteurs de transcription comme SP1 ou ERF1 pour la liaison sur leurs sites, et bloquant l'accès à la chromatine pour la machinerie de transcription (Budhram-Mahadeo *et al.*, 1999 ; Hoffman *et al.*, 2002) (**Figure 29A**)

de recruter de façon active des complexes capables de modifier la chromatine comme des HDACs ainsi que des complexes répresseurs tels que Sin3a ou HP1 (Nguyen *et al.*, 2005) (**Figure 29B**)

de séquestrer des facteurs activateurs comme le récepteur aux glucocorticoïdes (GR), l'interaction physique entre ces deux protéines abolissant ainsi les capacités activatrices du GR (Maiyar *et al.*, 1997) (**Figure 29C**)

d'inhiber l'activité transcriptionnelle d'un gène par l'intermédiaire d'ARNs non codant (ARNnc) de deux types: les micros ARNs (miARNs) qui se lient sur la séquence 3' non traduite d'un ARN et entraînant sa dégradation, ou *via* de long ARNs intergéniques non codant (lincARNs) (Rinn et Huarte, 2011) (**Figure 29D**).

F-Remarques générales sur l'activation et la répression transcriptionnelle

L'avènement des techniques d'analyse à grande échelle des régulations transcriptomiques et du recrutement des facteurs de transcription a permis de généraliser certaines observations et questionnements. Par exemple, il apparaît que la majorité des facteurs de transcription peuvent, de façon coordonnée, activer un ensemble de gènes et en réprimer d'autres, en recrutant, selon le contexte chromatinien, des co-facteurs activateurs ou au contraire répresseur de la transcription. Les mécanismes exacts qui sont à l'origine de ces recrutements différentiels et activités opposées ne sont pas pour le moment clairement établis. Plusieurs hypothèses ont été émises à partir des études sur p53, mettant en avant la nature de la séquence nucléotidique de liaison de la protéine comme la séquence consensus en elle-même (Wang *et al.*, 2009), l'espacement entre les 2 demi-sites constituant ses éléments consensus (Wang *et al.*, 2010; McKenzie *et al.*, 2010), ou encore l'orientation de la séquence consensus (Johnson *et al.*, 2001 ; Godar *et al.*, 2008).

Tableau 4: Classification des RNs basée sur l'arbre phylogénétique de cette superfamille (Escriva et al., 1997 ; Laudet, 1997).

Récepteur	Nomenclature	Ligand
TR α	NR1A1	Thyroid hormone
TR β	NR1A2	Thyroid hormone
RAR α	NR1B1	Retinoic acid
RAR β	NR1B2	Retinoic acid
RAR γ	NR1B3	Retinoic acid
PPAR α	NR1C1	Fatty acids, leukotrienne B fibrates
PPAR β	NR1C2	Fatty acids
PPAR γ	NR1C3	Fatty acids, prostaglandine J thiazolidinedione
Rev-erb α	NR1D1	Orphan
Rev-erb β	NR1D2	Orphan
ROR α	NR1F1	Cholesterol, cholesteryl sulfate
ROR β	NR1F2	Retinoic acid
ROR γ	NR1F3	Orphan
LXR α	NR1H3	Oxysterol, T0901317, GW3965
LXR β	NR1H2	Oxysterol, T0901317, GW3965
FXR α	NR1H4	Bile acids, fexaramine
FXR β	NR1H5	Lanosterol
VDR	NR1I1	Vitamin D, 1.25-dihydroxyvitamin D
PXR	NR1I2	Xenobiotics, 16a-cyanopregnenolone
CAR	NR1I3	Xenobiotics, phenobarbital
HNF4 α	NR2A1	Orphan
HNF4 γ	NR2A2	Orphan
RXR α	NR2B1	Retinoic acid
RXR β	NR2B2	Retinoic acid
RXR γ	NR2B3	Retinoic acid
TR2	NR2C1	Orphan
TR4	NR2C2	Orphan
TLL	NR2E2	Orphan
PNR	NR2E3	Orphan
COUP-TFI	NR2F1	Orphan
COUP-TFII	NR2F2	Orphan
EAR2	NR2F6	Orphan
ER α	NR3A1	Estradiol-17 β , tamoxifen, raloxifen
ER β	NR3A2	Estradiol-17 β , various synthetic compounds
ERR α	NR3B1	Orphan
ERR β	NR3B2	DES, 4-OH tamoxifen
ERR γ	NR2B3	DES, 4-OH tamoxifen
GR	NR3C1	Cortisol, dexamethasone, RU486
MR	NR3C2	Aldosterone, spiro lactone
PR	NR3C3	Progesterone, medroxyprogesterone acetate, RU486
AR	NR3C4	Testosterone, flutamide
NGF1-B	NR4A1	Orphan
NURR1	NR4A2	Orphan
NOR1	NR4A3	Orphan
SF1	NR5A1	Orphan
LRH-1	NR5A2	Orphan
GCNF	NR6A1	Orphan
DAX-1	NR0B1	Orphan
SHP	NR0B2	Orphan

IV-Les récepteurs nucléaires

Les récepteurs nucléaires (RN) constituent une super famille de facteurs de transcription ayant des rôles importants dans de multiples processus cellulaires et physiologiques tels que la croissance cellulaire, la différenciation, le maintien de l'homéostasie, la reproduction ou encore le développement. Ces fonctions des RN peuvent faire intervenir différents types de signalisation intracellulaire, l'une dite non génomique mettant en jeu des cascades de phosphorylation impliquant par exemple les MAPKs (Mitogen activated protein kinase), l'autre -générale à la superfamille- étant nommée voie génomique et impliquant la liaison directe ou indirecte de ces facteurs de transcription au niveau de la chromatine afin d'activer ou d'inhiber la transcription de leurs gènes cibles. Une des conséquences possibles des voies non-génomiques est également une modulation de l'activité transcriptionnelle de certains gènes.

A-La famille des récepteurs nucléaires

1-Classification

Il existe deux types de classifications des RNs. La première dite "classique" est basée sur leurs propriétés fonctionnelles, les discriminant ainsi au sein de 3 groupes (Parker, 1993; Tsai et O'Malley, 1994):

- la classe 1 regroupe les RNs aux hormones stéroïdes capable de se lier à l'ADN sous forme d'homodimère au niveau de palindromes inversés. Ce groupe correspond aux récepteurs aux œstrogènes (ER), androgènes (AR), progestagènes (PR), glucocorticoïdes (GR) et mineralocorticoïdes (MR).

- la classe 2: les récepteurs appartenant à cette classe ne présentent pas de spécificité en terme d'organisation de séquences liées même s'ils se lient préférentiellement à des demi-sites organisés en répétitions directes. Ils se fixent principalement à l'ADN sous forme d'hétérodimères avec RXR (9-cis retinoic acid retinoid X receptor) qui est considéré comme partenaire privilégié ou même obligatoire de la plupart des RNs de cette classe dont ceux aux hormones thyroïdiennes (TR), à la vitamine D (VDR), aux acides rétinoïques all-trans (RAR) et PPAR (peroxisome proliferator activated receptor).

- la classe 3 réunit les récepteurs appelés orphelins c'est à dire les récepteurs n'ayant pas de ligands ou ceux dont le ligand n'a pu être identifié. Elle inclut entre autres les COUP-TFs (Chicken ovalbumin upstream transcription factors), ERRs (ER-related receptors)...

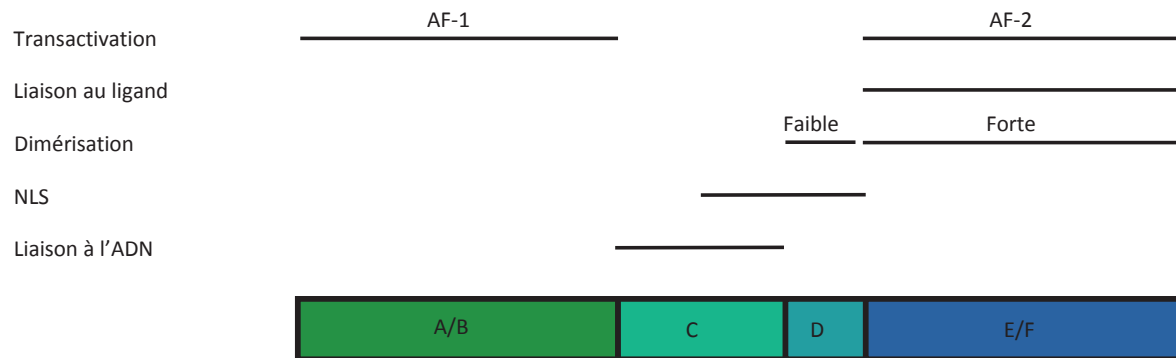


Figure 30: Structure des récepteurs nucléaires (Adapté de Tsai et O'Malley, 1994).

Les récepteurs nucléaires sont organisés en 4 modules, nommés A/B, C, D et E/F, chacun de ces domaines portant une activité spécifique mentionnée au dessus.

AF: «Activation Function» ; NLS: «Nuclear Localization Signal».

La seconde classification des RNs se base quant à elle sur l'arbre phylogénétique des membres de cette superfamille, qui a permis l'identification de 6 sous-groupes (Escriva *et al.*, 1997; Laudet, 1997) (**Tableau 4**).

2-Structure

La structure des RNs se base de façon générale sur une architecture commune, dite modulaire, qui s'organise autour de 4 domaines: A/B, C, D et E/F (**Figure 30**)

- **le domaine A/B**, également appelé domaine modulateur présente une certaine variabilité en terme de taille et de séquence entre les différents membres de la superfamille. Cette variabilité est accrue pour certains RNs par la présence d'exons alternatifs, de plusieurs promoteurs,... qui permettent l'expression de plusieurs isoformes d'un même récepteur. Ce domaine contient chez la plupart des RNs une fonction de transactivation, nommé AF-1 (Beato *et al.*, 1995) qui peut être activée *via* des événements de phosphorylation par les MAPK par exemple dans le cas des récepteurs ER α et ER β (Kato *et al.*, 1995 ; Bunone *et al.*, 1996 ; Trembaly *et al.*, 1997). Dans sa partie la plus C-terminale, ce domaine peut également moduler la liaison du RN à l'ADN, en influant sur sa capacité à interagir avec d'autres membres de cette super famille par exemple (Zechel *et al.*, 1994 a et b).

- **le domaine C** correspond au domaine de liaison à l'ADN, ou DBD qui est le plus conservé des domaines des RNs, présentant de 40 à 50 % d'identité entre les membres de la superfamille, et constituant de fait sa signature phylogénétique (Evans, 1988). La fonction de ce domaine est de permettre la liaison des RNs à l'ADN sous forme de monomère, homodimère ou hétérodimère (Leid *et al.*, 1992a) sur des éléments de réponse généralement organisés en répétitions de 2 demi-sites dont l'orientation et l'espacement varie (Gronemeyer et Laudet, 1995 ; Mangelsdorf *et al.*, 1995). Sa structure de base s'articule autour de deux doigts de zinc de type C₂C₂ qui permettent l'exposition de deux hélices alpha appelées boîtes P et D. La première permet la reconnaissance spécifique des séquences des demi-sites, tandis que la seconde est impliquée quant à elle dans la dimérisation du RN et la discrimination entre l'espacement et l'orientation des deux demi-sites.

- **le domaine D**, ou région charnière, est une région dont la séquence et la taille sont très variables. Ce module serait dans la plupart des cas très flexible, étant potentiellement capable d'effectuer une rotation de 180° dans le but d'aider la liaison du récepteur sur l'ADN (Glass, 1994). Cette région est également importante pour l'hétérodimérisation du récepteur

(Wilson *et al.*, 1992 ; Lee *et al.*, 1993) et peut contenir une surface d'interaction pour des protéines de type co-répresseur comme dans le cas du TR (Chen et Evans, 1995 ; Horlein *et al.*, 1995).

- **le domaine E/F** porte plusieurs fonctions, dont celle de liaison au le ligand (LBD) dans le cas des RNs non orphelins, mais permet également la dimérisation du récepteur, son interaction avec des protéines de choc thermique (HSPs), et possède des domaines impliqués dans la localisation nucléaire de la protéine. D'un point de vue structural, le LBD des RNs est constitué de 12 à 13 hélices α et d'un ou deux feuillets β (Bourguet *et al.*, 1995 ; Renaud *et al.*, 1995 ; Wagner *et al.*, 1995; Bourguet *et al.*, 2000). La liaison du ligand s'opère au sein d'une poche hydrophobe, repliée entre 3 hélices α organisées de façon anti-parallèle comprenant un coeur central (H5/H6 et H9/H10) entouré de deux couches d'hélices α formées de H1-H4 et de H7/H8 et H11. Enfin, ce domaine présente une fonction de transactivation ligand dépendante nommée AF-2 (Gronemeyer et Laudet, 1995), portée par l'H12. Cette région permet un recrutement direct de cofacteurs par le RN, qu'ils soient activateurs ou répresseurs de la transcription.

Sept récepteurs orphelins ne présentent pas cette structuration modulaire classique. Il s'agit du groupe KNI (avec le récepteur chez Drosophile KNI, KNRL et EGON), ODR-7 présent chez *C.elegans*, DAX-1 et SHP-1 présents chez l'Homme et enfin Trithorax chez la Drosophile (Laudet, 1997).

3-Mécanisme d'action

a-La localisation

Les récepteurs nucléaires sont des protéines majoritairement nucléaires où elles régulent l'expression de leurs gènes cibles. Cependant, en absence de ligand, certains RNs, et en particulier des récepteurs aux stéroïdes comme le GR, peuvent être retrouvés dans le cytoplasme en interaction avec des protéines chaperons les HSPs (Heat Shock Proteins, hsp90, hsp70 et hsp 56) (Pratt et Toft, 1997; Tsai et O'Malley, 1994). La liaison du ligand dans ce cas est nécessaire pour la relocalisation des récepteurs au sein du noyau. En absence de ligand, les RNs localisés dans le noyau peuvent favoriser la mise en place des facteurs nécessaires à l'assemblage du CPI (McKenna et O'Malley, 2002), ou au contraire la capacité de recruter des co-répresseurs (Horwitz *et al.*, 1996; Love *et al.*, 2000).

b-La liaison du ligand

La liaison du ligand au sein du LBD provoque de nombreux changements conformationnels de la protéine au niveau de cette région. C'est ce remodelage structural qui est à l'origine de la dissociation des RNs aux HSPs (Allan *et al.*, 1992 ; Vegeto *et al.*, 1992 ; Weigel *et al.*, 1992) et qui permet ou qui stabilise leur dimérisation. Les études cristallographiques de cette région des RNs ont permis de montrer que la position de l'H12 dans l'espace était un élément structural critique pour la fonction des RNs après fixation d'un ligand. Par exemple, si ce ligand est agoniste, l'H12 bascule pour former une sorte de couvercle au dessus de la poche hydrophobe renfermant le ligand, sans établir de contact direct. Ce positionnement précis permet la dissociation des RNs aux co-répresseurs et induit la compétence de l'AF2 en créant une surface d'interaction pour des coactivateurs de la transcription (Brzozowski *et al.*, 1997; Shiau *et al.*, 1998). Cette surface est en particulier reconnue par des protéines présentant un/des motifs de type LXXLL (avec L= Leucine et X, n'importe quel AA) tels que TIF-2/GRIP-1 (Hong *et al.*, 1996 ; Voegel *et al.*, 1996), SRC-1/P160 (Onate *et al.*, 1995 ; Kamei *et al.*, 1996 ; Yao *et al.*, 1996) ou encore CBP/P300 (Chakravarti *et al.*, 1996 ; Hanstein *et al.*, 1996 ; Kamei *et al.*, 1996 ; Yao *et al.*, 1996). Au contraire, l'interaction du RN avec un antagoniste entraîne une rotation de l'H12 créant une conformation de l'AF-2 qui n'est pas propice à une interaction avec les coactivateurs (Brzozowski *et al.*, 1997).

c-Dimérisation et liaison à l'ADN

Le changement conformationnel induit par la liaison au ligand permet aux RNs de s'homo- ou s'hétérodimériser avec un autre RN; ou encore de stabiliser les dimères préformés en absence de ligand. Le récepteur RXR est le partenaire privilégié des RNs lors des phénomènes d'hétérodimérisation, ce qui lui confère une importance cruciale pour de nombreux processus biologiques. Ces phénomènes d'hétérodimérisation peuvent être classés en deux types. On parle effectivement d'hétérodimérisation non permissive dans le cas où seul le ligand du partenaire de RXR peut activer le couple protéique (cas d'hétérodimères tels que RAR/RXR ou encore de VDR/RXR), et d'hétérodimérisation permissive lorsque ces hétérodimères peuvent être activés par le ligand de l'un ou de l'autre récepteur, comme dans le cas du PPAR/RXR. (Kliwer *et al.*, 1992 ; Willy *et al.*, 1995 ; Janowski *et al.*, 1996 ; Willy et Mangelsdorf, 1997).

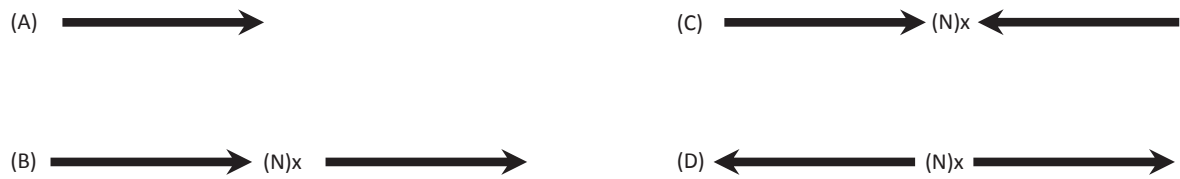


Figure 31: Liaison des récepteurs nucléaires à l'ADN (Adapté de Guigere, 1999).

Illustration des différentes organisations des sites reconnus par les récepteurs nucléaires.

(A) récepteur sous forme de monomère.

(B) répétition directe.

(C) palindrome.

(D) palindrome inversé.

N correspond au nombre de nucléotides séparant les deux demi-sites de liaison et qui permet, entre autre, la spécificité de liaison des RNs.

La structure spécifique des DBD des RNs permet leur ciblage sur des éléments de réponse composés de deux demi-sites de 6 paires de base. Ces séquences peuvent être organisées en palindromes ou répétitions directs ou inversés (Gronemeyer et Laudet, 1995), les deux 1/2 sites pouvant être séparés par un nombre de nucléotides variable (Umesono *et al.*, 1991) (**Figure 31**). Elles ont pour séquences consensus RGGTCA pour les récepteurs de type ER, RXR, COUP-TF et TR ou RGAACA pour les récepteurs aux stéroïdes tels que AR, PR ou encore GR (Cooney *et al.*, 1992 ; Aranda et Pascual, 2001). La discrimination entre ces deux types de consensus est opérée par trois acides aminés présents au sein de la boîte P du DBD. Ce sont ces trois paramètres (séquence, orientation et espacement) qui conditionnent la spécificité de liaison de RNs. Quelques récepteurs orphelins ont également la propriété de se lier à l'ADN sous forme de monomères, cette liaison pouvant être stabilisée par une interaction entre des séquences riches en A/T présentes en amont du demi-site et un domaine CTE (C terminal extension) porté par la région centrale (domaines C/D) de ces RNs (Wilson *et al.*, 1991 ; Sem *et al.*, 1997).

La régulation transcriptionnelle des gènes cibles des RNs peut également s'opérer par un recrutement indirect de ces protéines au niveau de l'ADN. Ceci implique l'établissement d'interactions physiques entre le RN et d'autres TFs (Pfahl, 1993 ; Karin, 1998), comme dans le cas du recrutement du ER sur des sites AP1 qui sont des séquences cibles pour les facteurs jun et fos (Webb *et al.*, 1999) ou bien sur les sites Sp1 (Safe, 2001, Kim *et al.*, 2003).

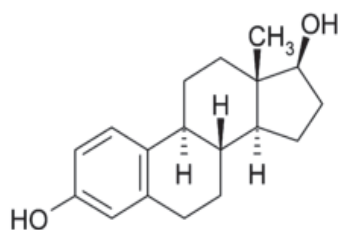
d-Activation des gènes par les RNs:

L'activation des gènes par les RNs peut se faire sur un promoteur minimal, suggérant une capacité des récepteurs à interagir directement avec la machinerie de transcription (Allan *et al.*, 1991). Des études menées *in vitro* ont effectivement montré que certains récepteurs aux stéroïdes activaient la transcription en stabilisant la formation du CPI au niveau des promoteurs cibles (Elliston *et al.*, 1990 ; Klein-Hitpass *et al.*, 1990 ; Tsai *et al.*, 1990) et cela *via* une interaction avec TFIIB (Ing *et al.*, 1992) ou la TBP (Sadovsky *et al.*, 1995).

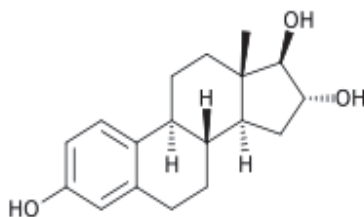
e-Inhibition des gènes par les RNs:

Contrairement à ce qui a longtemps été pensé, certains récepteurs ont la possibilité de lier l'ADN en absence de ligand et c'est cette liaison qui serait à l'origine d'une diminution du niveau basal de la transcription de certains gènes (Baniahmad *et al.*, 1992 a et b ; Cooney *et al.*, 1993). Le domaine des RNs impliqué serait le domaine E/F au niveau duquel on trouve

17 β -œstradiol (E2)



œstriol (E3)



œstrone (E1)

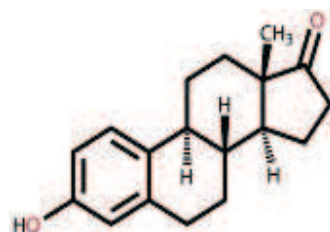


Figure 32: Formules développées des trois œstrogènes naturels chez la femme.

une région importante dans l'inhibition des gènes (Baniahmad *et al.*, 1992a) qui, comme pour TR par exemple, interviendrait en inhibant la formation du CPI toujours *via* une interaction avec TFIIB (Baniahmad *et al.*, 1993, Fondell *et al.*, 1993). Des études récentes ont également mis en avant l'existence d'éléments de réponse négatifs, comme pour GR, pour qui les RNs auraient une grande affinité. La liaison des récepteurs sur ces séquences entraînerait un changement de conformation, les rendant inaptes à activer la transcription (Tsai et O'Malley, 1994).

f-La voie non génomique

Certains RNs ont la capacité d'induire des effets cellulaires très rapides *via* une voie nommée non génomique. Par définition, les effets non génomiques d'hormones telles que les stéroïdes pourraient correspondre à des effets qui, suite à une stimulation hormonale, ont lieu rapidement et surtout ne sont pas bloqués par des inhibiteurs de la transcription ou de la synthèse protéique comme l'actinomycine D ou le cycloheximide (Wehling, 1997). Les effets non génomiques ont deux mécanismes d'intervention: ils peuvent soit interagir avec un récepteur cible présent à la membrane plasmique, soit interagir avec des protéines non spécifiques ou bien des lipides membranaires.

B-Le cas du récepteur aux œstrogènes: ER

1-Les œstrogènes

Les œstrogènes sont des hormones femelles dérivées du cholestérol ayant des rôles importants dans divers processus physiologiques et physiopathologiques comme la reproduction, les cancers du sein ou de l'endomètre, les maladies cardiovasculaires ou encore l'ostéoporose. Chez la femme, ils sont au nombre de trois: le 17β -œstradiol (E2) qui est majoritaire, l'œstriol (E3), présent uniquement pendant la gestation et l'œstrone (E1) (**Figure 32**). L'E2 est également présent chez l'Homme (Simpson *et al.*, 2002) mais en concentration très faible. Il existe également des composés ayant une analogie structurale et fonctionnelle avec les œstrogènes comme les phyto-œstrogènes qui sont des composés d'origine végétale ou bien des composés de synthèse, appelé modulateurs sélectifs des récepteurs aux œstrogènes (SERM) comme le tamoxifène (TAM), qui peuvent avoir des actions agonistes ou antagonistes par rapport aux récepteurs aux œstrogènes.

2-Les récepteurs aux œstrogènes

La majorité des effets des œstrogènes est exercée par deux récepteurs, ER α (nommé ici ER) et ER β , deux facteurs de transcription ligand-dépendants (Katznellenbogen et Katznellenbogen, 1996 ; Kuiper *et al.*, 1998).

a-ER α

Le récepteur ER est codé par le gène ESR1, présent sur le bras long du chromosome 6 chez l'Homme: 6q25.1 (Menasce *et al.*, 1993). L'expression de ce gène aboutit principalement à la synthèse d'une protéine de 66kDA (Reid *et al.*, 2002). Cependant, la découverte d'un épissage alternatif au niveau de l'exon 1 a permis d'identifier une forme tronquée de 46kDA où le domaine A/B est manquant (Flouriot *et al.*, 2000). Après liaison du ligand, ER *via* son domaine C, reconnaît son ERE présent sur l'ADN, ce qui lui permet, en relation avec d'autres protéines, de réguler ses cibles.

b-ER β

Le récepteur ER β , découvert en 1996, est codé par un gène autre que celui codant ER (Kuiper *et al.*, 1996 ; Mosselman *et al.*, 1996 ; Katznellenbogen et Korach, 1997). Ces deux récepteurs diffèrent principalement dans leur région A/B avec seulement 18% d'homologie ainsi que dans leur région E/F avec 56% d'homologie (Katznellenbogen *et al.*, 2000). Cette différence marquée au niveau des deux domaines de transactivation suggère que ces récepteurs ont une action différente. Si on prend comme exemple le domaine AF-2 qui permet une transactivation ligand-dépendante, il est important de noter que pour certains composés comme le tétrahydrochrysène (THC), il y a un effet sélectif, le THC étant agoniste pour ER et au contraire antagoniste pour ER β (Sun *et al.*, 1999).

3-Rôle de ER après liaison d'un agoniste

Avant les années 2000, il était communément admis que les TFs tels ER restaient liés à leurs éléments de réponse et activaient la transcription aussi longtemps que le signal perdurait (Hahn, 1998 ; Berk, 1999). Cependant des études effectuées sur la dynamique d'association d'ER sur des séquences cibles ont fait évoluer ce concept, mettant en avant un recrutement cyclique et combinatoire de ER et de ses cofacteurs (Shang *et al.*, 2000 ; Métivier *et al.*, 2003).

Après liaison du ligand, dissociation des HSPs et translocation dans le noyau, ER régule ses gènes cibles par liaison directe sur son ERE ou bien *via* une interaction avec d'autres facteurs

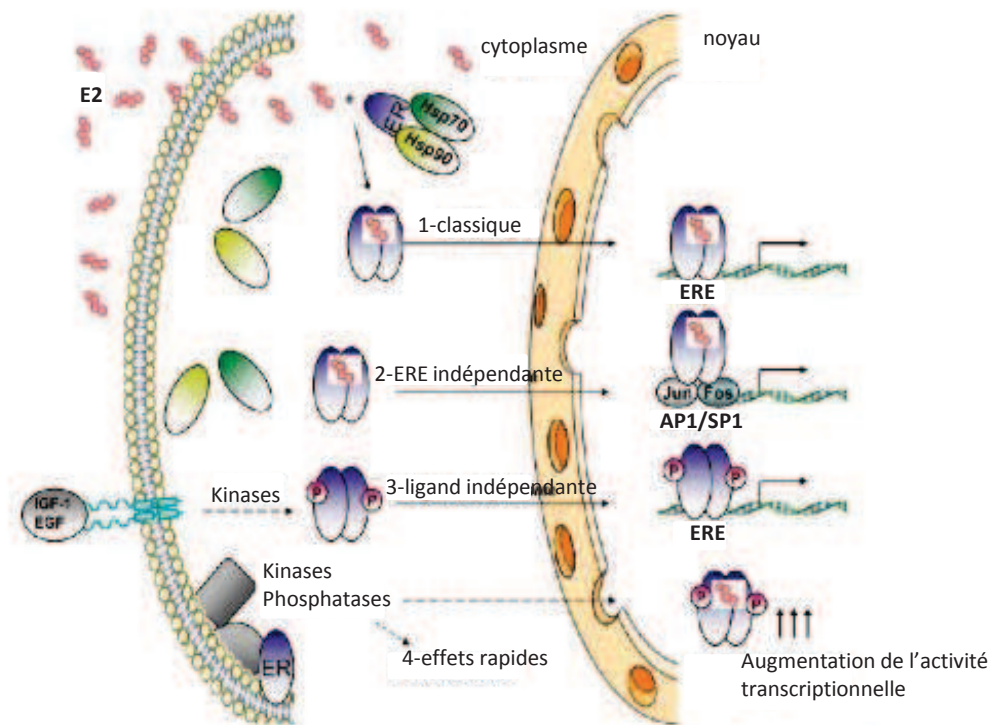


Figure 33: mécanismes d'action de ER (adapté de Reid et al., 2002 ; Heldring et al., 2007).

- 1- mécanisme d'action « classique »: après stimulation œstrogénique, ER se lie sur les ERE de ses gènes cibles.
- 2- mécanisme d'action ERE indépendante: ER se lie à l'ADN à travers des interactions protéines/protéines.
- 3- mécanisme d'action ligand-indépendant: ER est activé par des phosphorylations et se lie sur ses séquences cibles.
- 4- mécanisme d'action non génomique: des kinases sont activés par des récepteurs présents à la membrane et phosphorylent différentes protéines cytoplasmiques conduisant à des régulations transcriptionnelles de gènes cibles.

de transcription comme Sp1 ou AP-1 (Umayahara *et al.*, 1994 ; Li *et al.*, 2001) (**Figure 33**). Le long du génome, la liaison du ER peut se faire au niveau des promoteurs proximaux mais également, et ce majoritairement au niveau de séquences lointaines, de type enhancer (Carroll *et al.*, 2006). Cette liaison est favorisée par la présence de facteurs pionniers tels que FOXA1 dont le rôle serait de modifier la chromatine avant de la rendre abordable par les TFs. Une fois sur son site, ER recrute des coactivateurs comme CBP, P300 ou encore pCAF qui, d'une part possèdent une activité HAT leur permettant de modifier l'organisation de la chromatine (Shang *et al.*, 2000) et d'autre part permettent une connexion à la machinerie de transcription (Laudet, 1997).

Bien que la plupart des travaux initiés sur la régulation de la transcription médiée par ER concerne l'activation des gènes, il est important de noter qu'une grande partie (presque 50% selon les cellules et les temps d'activation) des cibles d'ER en présence du ligand sont réprimées. Il semblerait qu'AP-1, en plus de son rôle dans l'induction de la transcription par ER, possède une capacité d'action dans la répression de la transcription médiée par ER. En effet, une majorité des gènes concernés présentent un site de liaison AP-1. Cette répression nécessiterait également la présence d'un co-répresseur: NRIP1 qui pourrait agir avec ER dans un contexte AP-1 et cela dans le but de réprimer les gènes adjacents (Carroll *et al.*, 2006).

4-Rôle de ER après liaison d'un antagoniste

La liaison d'un antagoniste tel que le raloxifène (RAL) ou encore le TAM entraîne une conformation particulière de ER, et plus précisément de son domaine AF-2. Par exemple, en ce qui concerne le RAL, sa liaison sur ER entraîne une rotation de 130° de l'hélice H12 induisant une incapacité pour AF-2 de lier des co-activateurs (Brzozowski *et al.*, 1997). En ce qui concerne le TAM, il est connu pour entrer en compétition avec E2 dans la liaison à ER. Une fois le complexe TAM-ER formé, celui ci a la possibilité de recruter des co-répresseurs présentant la séquence conservée: LXXI/HIXXXI/L reconnue par AF-2.

5-ER et la voie non-génomique:

Il existe une action de l'E2 directement au niveau des membranes plasmiques et cela principalement à travers sa liaison à un ER membranaire ou à un autre récepteur membranaire, GPR30. La stimulation œstrogénique entraîne alors une rapide activation de voies de signalisation impliquant le calcium ou bien différentes protéines kinases (MAPK, PKA ou PKC), à travers leurs phosphorylation, et aboutit à des fonctions cellulaires très

importantes comme la prolifération cellulaire, la préservation du cytosquelette ou encore la survie en réponse à des dommages métaboliques (Migliaccion *et al.*, 1996 ; Kelly et Levin, 2001) (**Figure 33**). Ainsi, dans les cellules cancéreuses mammaires, l'activation œstrogénique de la cycline D1, qui permet la transition de la phase G1 à la phase S du cycle cellulaire, se réalise via la stimulation des kinases ERK ou PI3K (Phosphatidylinositol-3-kinase) (Razandi *et al.*, 2004).

V-Les facteurs pionniers

A-Définition

La chromatine condensée est un environnement peu propice pour le recrutement des TFs, pourtant nécessaire à l'établissement d'une régulation spatio-temporelle fine de la transcription. Depuis quelques années émerge la notion d'une classe particulière de facteurs de transcription caractérisés par une fonction commune qui serait de permettre le remodelage de la chromatine afin de la rendre compétente à la liaison d'autres TFs, et cela de façon indépendante de l'ATP : les facteurs pionniers (Cirillo *et al.*, 2002 ; Lupien *et al.*, 2008 ; Zareth *et al.*, 2008 ; Smale, 2010). Le recrutement de ces facteurs pionniers est dépendant de la séquence d'ADN, puisqu'ils présentent des affinités spécifiques pour certains types de motifs, mais également dépendant du contexte épigénétique et plus particulièrement de l'état de méthylation des queues des histones (Weigel et Jackle, 1990 ; Lupien *et al.*, 2008). Il semblerait également que les facteurs pionniers soient capables de fonctionner comme des marqueurs destinés à éviter un dépôt de marques histones liées à une répression transcriptionnelle comme H3K9me2 (Heintzman *et al.*, 2007 ; Lupien *et al.*, 2008 ; Heintzman *et al.*, 2009).

Ce chapitre est destiné à introduire les principaux facteurs pionniers permettant la liaison de ER à la chromatine. Il n'est pas exhaustif, car la définition de certains TFs comme étant des facteurs pionniers peut-être difficile. Par exemple, ER est capable de se lier à la chromatine de façon indirecte, *via* les TFs Jun/Fos ou SP1: doit-on considérer ces protéines comme étant des pionniers de ER? ER étant capable d'interagir de façon directe avec ces protéines dont l'action principale au niveau de la chromatine n'est pas forcément de permettre des modifications d'histones, ceci pourrait ne pas rentrer dans la définition d'un facteur pionnier. Mais certains des facteurs cités ci-après peuvent également remplir une de ces conditions. Deux caractéristiques certaines pour un facteur pionnier sont que sa présence est requise pour l'établissement d'un réseau de régulation, et que son recrutement doit précéder celui du/des facteur(s) pour lequel il agit en tant que pionnier.

B-Les facteurs FOX

Les facteurs FOX appartiennent à une famille protéique présentant un domaine de liaison à l'ADN appelé «winged-helix» en référence à sa conformation tri-dimensionnelle: trois hélices α de liaison à l'ADN entourées de deux boucles de polypeptides dont la forme figure

des «ailes» (Ramakrishnan *et al.*, 1993). Ce domaine est très conservé à l'intérieur de cette famille avec une homologie de 93% entre les protéines FOXA1, FOXA2 et FOXA3 (Lai *et al.*, 1991) ainsi qu'à travers les espèces (90% d'homologie avec la protéine Fork Head de la *Drosophile* (Cirillo *et al.*, 1998)). Il est important de noter que les histones de liaison, H1 et H5, dont le rôle est de rapprocher les nucléosomes afin de compacter la chromatine, ont cette même structure à une exception près, l'absence de la seconde «aile» (Ramakrishnan *et al.*, 1993 ; Cerf *et al.*, 1994). Cette similarité de structure entre ces protéines pose naturellement la question d'une redondance de fonction pour la liaison aux nucléosomes. En réalité, au contraire des histones de liaison, il semblerait que les protéines FOX aient la capacité à déplacer les nucléosomes afin de permettre un remodelage de la chromatine aboutissant à une structure plus ouverte (Cirillo *et al.*, 1998). Chez les mammifères, il existe plusieurs membres dans la famille des protéines FOX, chacun présentant une variabilité dans la séquence de son DBD, permettant ainsi la définition de classes allant de FOXA jusqu'à FOXQ (Katoh, 2004). Etant donné le contexte du sujet de cette thèse, deux membres de la classe FOXA ont une importance particulière.

1-FOXA1

FOXA1 est un TF principalement enrichi dans les cellules du foie, initialement caractérisé pour son rôle dans la liaison d'un enhancer du gène de l'albumine (Liu *et al.*, 1991 ; Hu *et al.*, 1992 ; Mc Pherson *et al.*, 1996). La particularité de FOXA1 est qu'il aurait *in vitro* la capacité à se lier de façon plus stable sur de la chromatine que sur de l'ADN nu (Cirillo et Zaret, 1999). Des études plus récentes ont également mis en avant son rôle *in vitro* dans le repositionnement des chromosomes (Zaret, 2002). Il pourrait également, du fait de sa structure, déplacer l'histone de liaison H1 *in vivo* (Cirillo *et al.*, 1998 ; Taube *et al.*, 2010) et permettre ainsi une ouverture locale de la chromatine (Cirillo *et al.*, 2002 ; Hurtado *et al.*, 2011).

Contrairement à beaucoup de TFs, la liaison de FOXA1 sur les nucléosomes ne serait pas facilitée par l'acétylation de résidus présents sur la queue des histones (Cirillo et Zaret, 1999) mais serait en partie due à une reconnaissance des marques H3K4me1/me2 (Lupien *et al.*, 2008 ; Eeckhoutte *et al.*, 2009), et nécessiterait une hypométhylation de l'ADN de ses sites de liaison (Serandour *et al.*, 2011). Il semblerait que d'autres protéines, non histones, comme CTCF pourraient être impliquées dans la liaison de FOXA1 sur la chromatine (Fu *et al.*, 2008 ; Zhang *et al.*, 2010 ; Hurtado *et al.*, 2011).

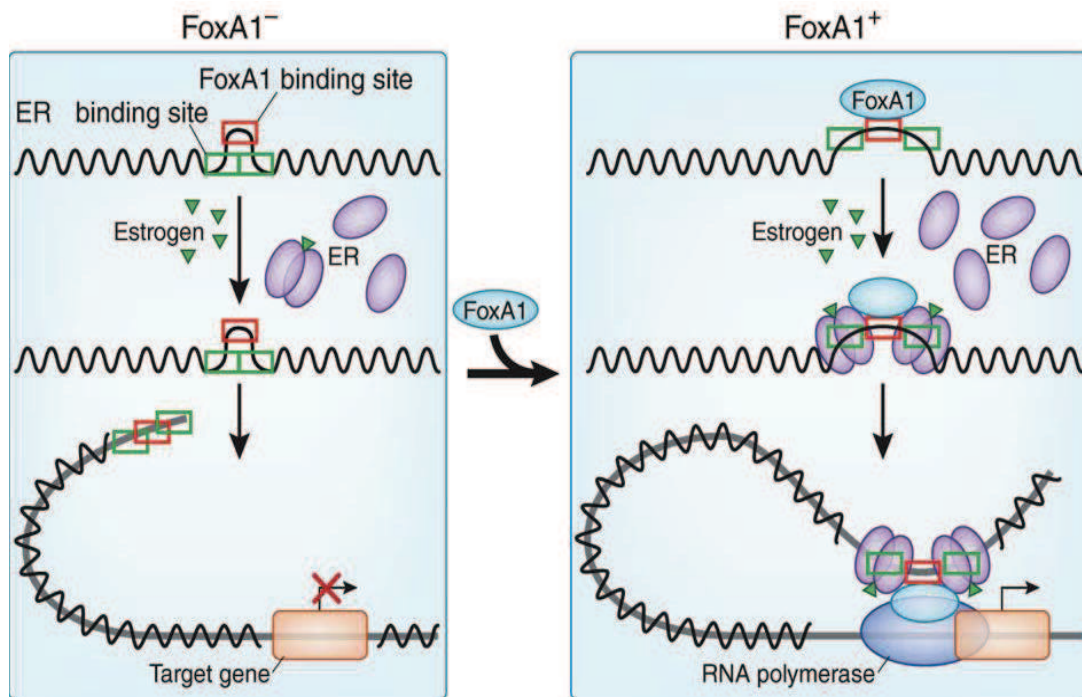


Figure 34: Rôle de FOXA1 dans la régulation transcriptionnelle médiée par ER (Cheung et Ruan, 2011).

En absence de FOXA1, les séquences cibles de ER sont inaccessibles au ER lié à l'E2, la structure chromatinienne locale étant condensée. Les gènes cibles de ER ne sont alors pas activés.

Au contraire, dans les cellules exprimant FOXA1, ce facteur pionnier se lie à ses sites de liaison où il permet une décompaction de la chromatine, rendant les séquences cibles de ER accessibles et donc une activation des gènes cibles de ER en présence d'E2.

FOXA1 a une importance particulière dans la régulation de la transcription médiée par ER. Initialement, des travaux ont en effet mis en avant, d'une part une corrélation entre l'expression des protéines ER et FOXA1 dans des cellules de cancer du sein et, d'autre part, une interaction (indirecte) entre ces deux protéines *via* la technique de double hybride chez la levure (Schoor *et al.*, 2001). De plus, l'importance de FOXA1 pour la régulation transcriptionnelle œstrogéno-dépendante avait déjà été déterminée sur des gènes spécifiques comme *TFF1* (Beck *et al.*, 1999) ou encore celui de la vitellogénine B1 (Robyr *et al.*, 2000). Des études plus récentes (Carroll *et al.*, 2005 ; Hurtado *et al.*, 2011) ont permis de généraliser ces travaux à l'échelle du génome en montrant l'importance de FOXA1 pour le recrutement de ER sur un grand nombre de ses sites de liaison (à hauteur de 50%) (**Figure 34**) mais également pour l'établissement de la régulation transcriptionnelle des gènes cibles de ER; une déplétion de FOXA1 par siRNA entraînant un défaut général de liaison d'ER et d'activation de la transcription de gènes comme *TFF1* ou encore *XBP-1*.

Le rôle pionnier de FOXA1 n'est cependant pas limité à ER. Des travaux sur le récepteur aux androgènes dans des cellules prostatiques (Gao *et al.*, 2003), ou d'autres portant sur le GR ont en effet montré son rôle dans la régulation de la transcription médiée par ces RNs (Holmqvist *et al.*, 2005).

La majeure partie du travail effectuée sur FOXA1 met en avant ses rôles pour la liaison d'autres protéines mais très rarement sa propre régulation. Il semblerait que ce TF pionnier soit régulé, *in vivo*, par l'acétylation de résidus portés par son DBD. Ces modifications post-traductionnelles diminueraient la liaison de FOXA1 au niveau des nucléosomes et donc incapaciteraient les mécanismes locaux de remodelage de la chromatine médiés par ce TF. Cependant, ces premiers résultats sont à prendre avec du recul, les niveaux de FOXA1 acétylés *in vivo* étant très faibles (Kohler et Cirillo, 2010).

2-FOXA2

FOXA2 est un autre membre de la famille des facteurs FOX, présentant une homologie de séquence importante avec FOXA1 (Lai *et al.*, 1991) ainsi qu'une certaine redondance de fonction (Lai *et al.*, 1990). Cette protéine est essentielle pour le développement du pancréas et du foie (Lee *et al.*, 2005 ; Gao *et al.*, 2008) et plus de 43% des gènes exprimés dans ces organes sont associés à FOXA2 (Wederell *et al.*, 2008). Tout comme FOXA1, FOXA2 a la capacité à réguler la transcription de certains gènes, par exemple *PDX1* (Gao *et al.*, 2008). En ce qui concerne les sites de liaisons de FOXA2, il semblerait qu'il y ait une

corrélation forte entre la présence de ce facteur et celle de H3K4me1 dans les nucléosomes flanquants son site de liaison (Hoffman *et al.*, 2010).

C-Les facteurs GATA

Les facteurs GATA font partie d'une grande famille de TFs identifiés initialement pour leur implication dans le développement des intestins. Les facteurs GATA-4, 5 et 6 sont ainsi exprimés dans les tissus dérivant de l'endoderme chez les mammifères (Arceci *et al.*, 1993 ; Laverriere *et al.*, 1994) où ils régulent l'expression de certains gènes (Bossard et Zaret, 1998 ; Gao *et al.*, 1998) comme celui de l'albumine (Jackson *et al.*, 1993 ; Bossard et Zaret, 1998). Au contraire, les facteurs GATA-1, 2 et 3 sont principalement impliqués dans l'hématopoïèse (Leonard *et al.*, 1993). Ces TFs ont plusieurs points communs dont la présence dans leur DBD d'un motif de type doigt de Zinc, présent en simple ou double exemplaires, qui leur permet de lier l'ADN sur des séquences caractéristiques de type «(A/T)GATA(A/G)» (Omichinski *et al.*, 1993 ; Pedone *et al.*, 1997). Ces facteurs sont capables de moduler la transcription de leurs gènes cibles *via* le recrutement de coactivateurs comme FOG ou encore CBP (Durocher *et al.*, 1997 ; Tsang *et al.*, 1997 ; Blobel *et al.*, 1998). Les facteurs GATA pourraient également, dans les cellules pluripotentes, entraîner une acétylation de la chromatine suite à leur liaison, permettant ainsi le recrutement d'autres TFs ou de la RNA-Pol II (Zaret, 1999).

Dans les cellules de cancer du sein, l'expression de GATA3 est corrélée à celle de ER (Lacroix et Leclercq, 2004) et des études plus récentes ont montrées que ces deux protéines pouvaient être présentes au niveau des mêmes régions génomiques, 45% des sites de liaison de ER dans les cellules cancéreuses mammaires MCF-7 recrutant également GATA3 (Kong *et al.*, 2011). Bien qu'une interaction physique entre GATA3 et ER n'ait pas été définie, l'hypothèse d'un rôle de GATA3 dans la signalisation œstrogénique a été confirmée ; en effet, la déplétion de GATA3 entraîne une baisse de l'expression de ER (liée à un déficit de recrutement de l'ARN-Pol II sur le gène ESR1) et de l'expression de gènes œstrogéno-dépendants comme SDF1 (Eeckhoutte *et al.*, 2007). De plus, d'autres travaux portant sur la régulation de la transcription par des facteurs GATA ont montré leur rôle dans l'établissement de boucles de chromatine au niveau du locus de la β -globine chez la Souris pour GATA1 (Vakoc *et al.*, 2005) ou encore au niveau du locus IL4/IL5/IL13 pour GATA3 (Spilianakis et Flavell, 2004).

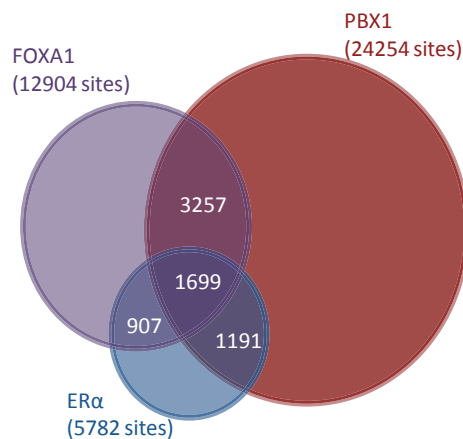


Figure 35: Recouvrement des cistromes de FOXA1, PBX1 et ER dans les cellules MCF-7 (Magnani et al., 2011).

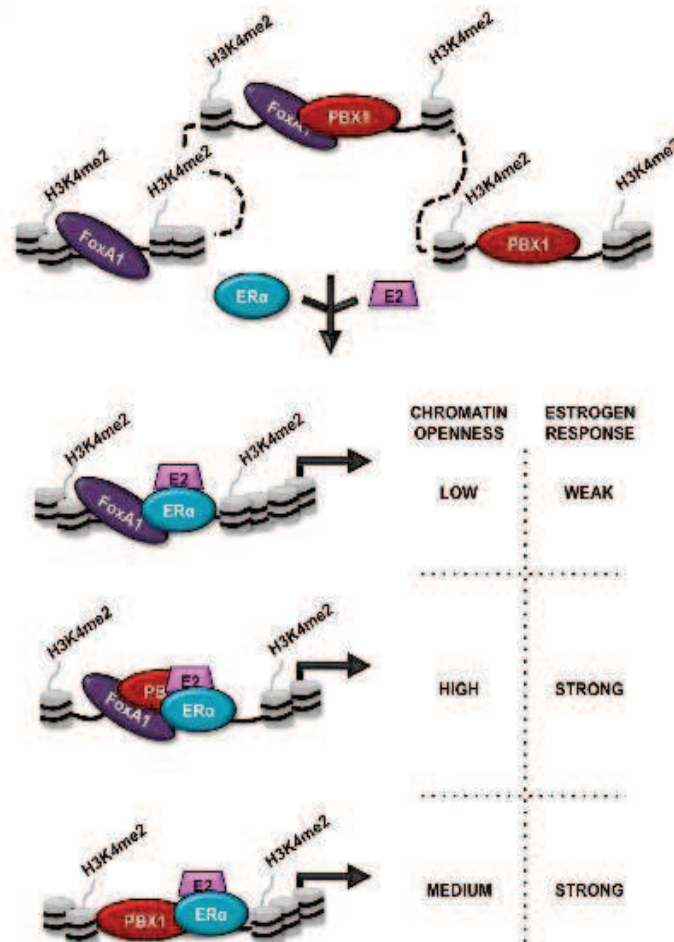


Figure 36: Illustration schématique du rôle de FOXA1 et de PBX1 dans la régulation transcriptionnelle médiée par ER (Magnani et al., 2011).

La présence de FOXA1 seul sur la chromatine entraîne une décompaction faible de la chromatine suivie d'une réponse transcriptionnelle aux œstrogènes basse. Au contraire, la liaison de PBX1 permet une ouverture plus importante de la chromatine au niveau de ses sites de liaison en relation avec une régulation transcriptionnelle médiée par ER forte. La présence de FOXA1 et de PBX1 sur la chromatine permet une ouverture importante de la chromatine et une réponse transcriptionnelle forte en réponse à une stimulation œstrogénique.

D-PBX1

PBX1 (Pre B cell leukemia homeobox-1) appartient à la famille des protéines à homéodomaine TALE (Three Amino acid Loop Extension) (Bertolino *et al.*, 1995). PBX1 est impliqué dans l'hématopoïèse, l'organogenèse (Di Martino *et al.*, 2001) ainsi que dans les cancers de la prostate, des ovaires ou de l'œsophage (Liu *et al.*, 2005 ; Kikugawa *et al.*, 2006 ; Park *et al.*, 2008 ; Yeh *et al.*, 2009). PBX1 a également la capacité à moduler l'activité transcriptionnelle de certains RNs comme le GR ou encore le TR (Subramaniam *et al.*, 1998 ; Wang *et al.*, 2001). Du fait de son expression importante dans les cellules de cancer du sein (Crijs *et al.*, 2007), des travaux récents se sont focalisés sur la caractérisation de son rôle potentiel dans la transcription E2-dépendante. Bien que l'action du ER semble principalement médiée par FOXA1 (Carroll *et al.*, 2005 ; Laganiere *et al.*, 2005 ; Carroll *et al.*, 2006 ; Eeckhoutte *et al.*, 2006 ; Lin *et al.*, 2007 ; Lupien *et al.*, 2008), il semblerait également que PBX1 y participe en tant que facteur pionnier (Magnani *et al.*, 2011). En effet, les sites de liaison du ER présentent un enrichissement important en motifs PBX1. De plus, des expériences de déplétion par siRNA ont montré que, dans les MCF-7, PBX1 a un rôle primordial dans la réponse aux œstrogènes, son absence empêchant la mobilisation de ER sur des sites où seul PBX1 est présent (i.e des sites ne recrutant pas FOXA1). Des expériences de ChIP et de FAIRE ("Formaldehyde Assisted Isolation of Regulatory Elements") ont de plus mis en avant la capacité de PBX1 à se lier à la chromatine avant la stimulation oestrogénique ainsi qu'à la remodeler en une structure moins compacte. Ceci illustre son rôle de facteur pionnier. Certains des sites de liaison de PBX1 recrutent également FOXA1 (**Figure 35**), ces deux facteurs auraient à ce niveau une action synergique pour le remodelage local de la chromatine et dans la régulation de la transcription E2-dépendante (**Figure 36**).

RÉSULTATS | CHAPITRE I

CONTEXTE DE L'ETUDE

En réponse à son environnement, constitué de signaux régulateurs endogènes et exogènes, une cellule doit être capable d'adapter son transcriptome. Au sein de l'équipe, nous travaillons sur les mécanismes impliqués dans la régulation spatio-temporelle de la transcription, et cela en particulier dans des contextes cellulaires œstrogéno-dépendants. Au début de ma thèse en 2009, des études sur la régulation du gène modèle *TFF1* avaient déjà été menées et avaient montré que le ER et ses cofacteurs étaient recrutés de façon dynamique sur le promoteur de ce gène (Shang *et al.*, 2000 ; Metivier *et al.*, 2003). Les caractérisations des transcriptomes œstrogéno-dépendants de cellules mammaires cancéreuses de type MCF-7 avaient permis de mettre en évidence le fait qu'une certaine proportion de gènes répondant aux œstrogènes étaient organisés en cluster, *i.e.* que des gènes colinéaires pouvaient présenter une réponse coordonnée à un même signal (Charpentier *et al.*, 2000; Ciciatello *et al.*, 2004). Les mécanismes à l'origine de telles régulations orchestrées par un même signal à l'échelle de plusieurs gènes œstrogéno-dépendants n'étaient par contre pas décrits en détail. Une des possibilités, peut-être un peu naïve, envisagée à l'époque était que, ER se liant préférentiellement sur des sites distants des gènes qu'il régule (Carroll *et al.*, 2006), l'organisation tridimensionnelle de la chromatine pouvait jouer un rôle important. Par exemple en rapprochant un site ER unique des différents promoteurs des gènes d'un cluster. Les études portant sur l'importance du complexe CTCF/cohésine (Nativio *et al.*, 2008 ; Wendt et Peters, 2009) dans la signalisation œstrogénique (Schmidt *et al.*, 2010); et les données de ChIA-PET (Fullwood *et al.*, 2009b) n'étaient à l'époque pas encore parues.

OBJECTIFS

L'objectif de cette étude était donc de comprendre les mécanismes spatio-temporels permettant à ER de contrôler un cluster de gènes. Pour cela, nous avons travaillé sur deux types cellulaires distincts, les MCF-7 qui expriment ER de façon endogène et les MDA::ER qui expriment ce récepteur de façon ectopique. Nous avons en effet décidé de nous placer dans des contextes présentant des transcriptomes œstrogéno-dépendants distincts, caractérisés au laboratoire avant mon arrivée. En particulier, la géométrie des régulations du cluster de gènes TFF incluant le gène modèle *TFF1* était différente entre les deux lignées. Notre hypothèse était que cette étude comparative allait permettre de définir des mécanismes cellule-dépendants initiés par ER à l'origine de ces différences.

Nous avons donc dans un premier temps caractérisé les sites de liaison de ER et de CTCF et RAD21 dans notre modèle cellulaire MDA::ER par des expériences de ChIP-chip, pour évaluer ensuite les mécanismes dynamiques mis en place par ER au niveau de ces sites. Nous avons en parallèle évalué les remodelages structuraux de cette région génomique, puis évalué l'importance d'autres facteurs de transcription impliqués dans cette régulation.

CONTRIBUTIONS AUX DIFFERENTS RESULTATS

- Réalisation des expériences de DNA FISH, sans siRNA ainsi que dans un contexte de déplétion de RAD21, et leurs analyses
- Réalisation des expériences de ChIP, ER, CTCF et RAD21, GATA2 et 3, H3K4me1/me2, H3K27ac, H3K14ac, H3R17me2
- Evaluations des effets des TFOs sur le recrutement de ER (ChIPs) et sur l'expression des gènes du cluster (RT-qPCRs)
- RT-qPCR et ChIPs après déplétions de CTCF, RAD21, GATA3 et de FOXA1 dans les MCF-7 et après GATA2 et GATA3 dans les MDA::ER
- les expériences de FAIRE
- Les expériences de MeDIP

CELL-SPECIFIC AND SPATIO-TEMPORAL CONTROLS OF THE ESTROGEN-RESPONSIVE TREFOIL FACTOR (TFF) LOCUS ACTIVITY

Justine Quintin¹, Christine Le Péron¹, Gaëlle Palierne¹, Maud Bizot¹, Stéphanie Cunha^{1,*}, Aurélien A. Sérandour^{1,**}, Stéphane Avner¹, Frédéric Percevault¹, Sébastien Huet¹, Erwan Watrin³, Jérôme Eeckhoutte^{1,4}, Vincent Legagneux², Gilles Salbert¹, and Raphaël Métivier¹

¹ Equipe SP@RTE, UMR CNRS 6290. Equipe labellisée Ligue contre le Cancer Université de Rennes I. Campus de Beaulieu. 35042 RENNES CEDEX. FRANCE. ² Equipe EGD, UMR CNRS 6290. Université de Rennes I. Faculté de Médecine. 2 avenue du Professeur Léon Bernard. 35043 RENNES CEDEX. FRANCE. ³ Equipe CC, UMR CNRS 6290. Université de Rennes I. Faculté de Médecine. 2 avenue du Professeur Léon Bernard. 35043 RENNES CEDEX. FRANCE. ⁴ INSERM U1011, Université Lille-Nord de France, Faculté de Médecine de Lille-Pôle Recherche. 59045 Lille. FRANCE.

Contact:

Raphaël Métivier, Equipe SP@RTE, UMR CNRS 6290. Batiment 13. Université de Rennes 1. Campus de Beaulieu. 35042 Rennes Cedex. Tel: +33 (0)2 2323 5142 Fax: +33 (0)2 2323 6794; E-mail: raphael.metivier@univ-rennes1.fr

Additional footnotes

* Present Address: Huntsman Cancer Institute. Salt Lake City, Utah. USA.

** Present Address: CRI, Cancer Research UK, Cambridge. UK.

Highlights

- cell-specific transcriptional regulations of a genomic region
- a dynamic interactome organizes these regulations in space and time
- functional characterization of ER interactome using TFOs to interfere with ER binding
- GATA proteins restrict ER actions in the absence of estradiol

Short Title

ER-mediated regulations of the TFF cluster in 4D

Keywords

Estrogen receptor; Transcription; Interactome; Triplex forming oligonucleotides; GATA; FOXA1.

SUMMARY

Cell-specific transcriptional regulations exerted by the estrogen (E2) receptor alpha (ER) heavily rely upon timely and spatially coordinated processes. We engaged a comparative analysis of such dynamic molecular events at the TFF locus harbouring a cluster of genes co-regulated by E2, in two distinct breast cancer cell lines. Using a combination of methods, we show that the recruitment of ER on cell-specific sites triggers dynamic local modifications of chromatin, which are coordinated in time all along the locus. DNA-FISH experiments further demonstrate that these changes are associated with an E2-dependent reduction in plasticity of this genomic region and are dependent upon cohesin. Importantly, 3C/4C experiments and the use of triplex forming oligonucleotides (TFOs) allowed us to precisely map the three-dimensional network of regulatory events that permits the estrogenic response of this genomic region. These data also evidenced an unexpected functional redundancy of enhancers towards the regulation of a gene. Finally, GATA factors were found to be involved in the cell-specific regulations of the TFF cluster by controlling ER actions in the absence of E2.

INTRODUCTION

Signaling pathways initiated by steroid hormones such as 17 β -estradiol (E2) provide model systems to study complex and highly organized transcriptional regulations in mammalian cells. Hence, exposure to estrogens leads to dynamic and precise modulations of the transcription status of cell-specific repertoires of genes mediated through specific ligand-dependent transcription factors, termed Estrogen Receptors (ER α -ER throughout the manuscript- and ER β ; McEwan, 2009). On model gene promoters, such as *TFF1*, ER together with a number of its cofactors associate with cognate binding sites in a cyclic manner to direct their transcription (Metivier et al., 2003; Shang et al., 2000). These dynamic mechanisms by which ER regulates transcription require the integration of several parameters: DNA sequence (*cis* elements), chromatin structure and epigenetic information, the action of factors in *trans* (transcription factors, cofactors and Pol II machinery) and the spatial organization of the genome within the nucleus (Metivier et al., 2006; Welboren et al., 2011). The three-dimensional organization of the genome within chromatin has indeed been proven to constitute the basis of its coordinated expression during cell differentiation or response to stimuli (Hurst et al., 2004; Perkins et al., 2004, van Driel et al., 2003). This is notably illustrated by the existence of clusters of co-regulated genes that can originate from genetic and epigenetic information or from chromatin dynamics in itself. Such a coordinated regulation of co-linear genes can even in some instances be dependent upon one global regulating unit, as occurs at β -globin, Mrf4 or Hox clusters (Chambeyron and Bickmore, 2004; Li et al., 2006; Carvajal et al., 2008; Tang et al., 2008).

Genome-wide analyses of ER chromatin binding sites (BS) have demonstrated that ER rarely binds to the proximal promoter of its target genes. Rather, ER is mobilized onto intergenic and intronic sequences (Carroll and Brown, 2006; Wellboren et al., 2009) that may communicate with putative target genes *via* long-distance intrachromosomal interactions (Fullwood et al., 2009). Whether these distant elements are acting as global regulators for clustered E2-responsive genes, such as those described in Cicatiello et al. (2004), is still an intriguing question. Another crucial outcome of these genome-wide studies was the demonstration that other transcription factors were responsible for the accurate targeting of ER onto the whole genome (Zaret and Carroll, 2011). These factors include FOXA1 (Carroll et al., 2006; Hurtado et al., 2011), AP2 (Tan et al., 2011), or the recently characterized PBX1 (Magnani et al., 2011). FOXA1 is considered as a pioneer factor for it may act as an allosteric sensor for histone marks associated with active or poised chromatin such as H3K4 mono/di-methylation and further prepare chromatin for subsequent binding of ER (Lupien et al., 2008; Motalebijpour et al., 2008; Serandour et al., 2011). The roles of these pioneer factors in processes governing the regulation of clustered genes are yet not precisely described.

We engaged here a thorough analysis of mechanisms involved in the coordination of the estrogenic response of collinear genes by focusing on one cluster of E2-sensitive genes. A systematic comparison of results obtained in MCF-7 cells with those arising from experiments performed in ER- and FOXA1- negative MDA-MB231 cells engineered to constitutively express ER (MDA::ER; **Figure S1**) allowed us *i)* to interrogate whether the introduction of ER in these cells recapitulate main regulations observed in MCF-7 at the TFF locus and *ii)* to define key parameters and factors specifying the transcriptional output at E2-regulated clustered genes.

RESULTS

E2 provokes dynamic transcriptional regulations and a global spatial reorganization of the TFF cluster. To consider the processes governing the co-regulation of collinear genes by E2, we first characterized MDA::ER and MCF-7 repertoires of E2-sensitive genes (GSE23850). This led us to identify in both cell lineages clusters of E2-regulated genes, defined as regions comprising at least 3 regulated genes within a window of 7 genes (**Figure S2**). As a paradigm, we focused here on one of such clusters located within locus 21q22.3 and termed TFF (**Figure 1A**). This cluster includes the prototypical estrogen-responsive gene *TFF1* and was previously identified in MCF-7 cells (Chinery et al., 1996). This locus contains a different number of E2-sensitive genes in MCF-7 and MDA::ER cells (**Figure 1B** and **Figure S3**). Results obtained by RT-qPCR assays confirmed *i*) the timely coordinated regulation of these genes by E2 (**Figure 1C**) and *ii*) that ER is the main transcription factor responsible for their E2-responsiveness, as indicated by their non-responsiveness to E2 following a 24h pre-treatment of the cells with the antiestrogen ICI that abrogates ER actions (**Figure 1C**). The analysis of the kinetics of the RNA Polymerase II (PolII) recruitment on the promoter regions of E2-sensitive genes by ChIP-qPCR further indicated that their transcriptional modulation was cyclical in nature, in both cell types (**Figure 1D**). These dynamic processes were specific for E2-regulated genes, as controlled by the constant engagement of PolII on the non-regulated *PDE9A* promoter in MDA::ER cells.

We next envisioned that these coordinated regulations may involve an E2-dependent spatial reorganization of this genomic region. This hypothesis was confronted to DNA-FISH experiments performed with a mix of fluorescent probes generated from BACs covering a large part of the TFF region (**Figure 1A**). Noteworthy, MDA::ER cells contain three TFF loci whereas MCF-7 presented a higher and heterogeneous number of loci (from 8 to 12). As this variability hindered the correct evaluation of the experimental results, we focused on MDA::ER cells in which we evidenced a clear compaction of the chromatin domain containing the TFF cluster after a 50 min treatment with E2 (**Figure 1E**). This could be assigned to a smaller dispersion of the distances separating paired central (RP11-113F1) and 5' (RP11-814F13) probes following exposure to E2 (**Figure 1F**). Importantly, these changes were not correlated with a significant reduction of the nuclei volume (**Figure S4**). Interestingly, whilst there were no changes in the distribution of distances between probes generated using the central and the 3' BACs (CTD-1033M14), the distribution of the distances separating probes located at both extremities of the TFF region were again significantly different. Furthermore, quantile-quantile representations of data (Q-Q plots, **Figure 1F**) indicated that observed changes mostly reflect a disappearance of large distances separating paired probes. The TFF locus thus undergoes spatial rearrangements under E2 treatment in MDA::ER cells, reflected by a more constrained three-dimensional conformation. ER has been proven to modulate the frequencies of interactions between distant ER binding sites (ERBS) and promoters of either up- or down-regulated genes (Fullwood et al., 2009; Hsu et al., 2010). This led us to question next whether a particular or a limited subset of ERBS within the TFF cluster could orchestrate the observed coordinated behaviors through long-range interactions.

Cell-specific ER cistrome and geometry of the TFF cluster. To consider the abovementioned hypothesis, we first had to identify MDA::ER ERBS. We thus ran ER ChIP experiments on chromatin prepared from MDA::ER

cells treated for 50 min with 10^{-8} M E2, with resulting samples hybridized on custom DNA arrays designed to cover genomic regions containing E2-sensitive clusters (**Figure S2**). These assays identified 17 ERBS at a FDR<5% within the TFF cluster (**Figure 2A**). Consistent with an overall low overlap between MDA::ER and MCF-7 ERBS on the genomic regions spotted on the arrays (**Figure S5**; MCF-7 data from Carroll *et al.*, 2006), only three of these binding sites were common to both cell lines (in green in **Figure 2A**) including the comBS2 located within the *TFF1* promoter. The cell-specificity of these ERBS were controlled by independent ChIP-qPCR experiments (**Figure 6**, and their inherent characteristics of enhancers verified by transfection experiments, cell-specific enrichment in canonical histone marks for active or poised enhancers and also in Pol II according to recent data showing that transcription can occur at enhancers (Hah *et al.*, 2011) (**Figure S6**).

We next used 4C-qPCR to characterize the spatial proximity of these ERBS with the promoters of E2-regulated genes. We choose 4C over 3C to get enough material of the linked targets for a thorough screening with high sensitivity. These assays were conducted on chromatin prepared from MCF-7, MDA::ER as well as MDA-MB231 cells treated for 50 min with E2. Results of these experiments (**Figures S7 and S8**) are schematized in **Figure 2A** and summarized within **Table S1**. These 4C assays recovered all but one of the interactions previously determined by ChiA-PET (Fullwood *et al.*, 2009) as linking ERBS to regulated promoters in MCF-7 cells, and further identified 38 more (**Figure 2B**). In accordance with DNA-FISH results, these experiments indicated that major spatial constraints of this chromatin domain might involve interactions between central and distant regions mainly located 5' of the TFF cluster (approx. 70% of the ERBS-promoters interactions) in both MCF-7 and MDA::ER cells. Interestingly, the *RIPK4* promoter was situated in the spatial vicinity of more ERBS in MDA::ER than in MCF-7 cells (**Figure 2B and 2C**). This might be correlated with the estrogenic regulation of this gene in the former cell line but not in the latter (**Figure 2C**). Importantly, 27% of the interactions detected in MDA::ER were also detected in ER negative MDA-MB231. This indicates *i*) that some of the MDA::ER ERBS may be pre-existing enhancers recruiting other transcription factors regulating the activity of all considered genes but *TFF3*; and *ii*) that the expression of ER in this system is sufficient, either directly or not, to remodel the spatial organization of this genomic region since 73% of the interactions detected in MDA::ER were not present in MDA-MB231. As illustrated within **Figure 2D and 2E**, the MCF-7 and MDA::ER network of chromatin loops between ERBS and promoters of regulated genes is complex and cell-specific. One interesting observation is that there are more singleton interactions in MCF-7 than in MDA::ER. Hence, while there was 6 ERBSs interacting with 1 gene in MCF-7, there was only one of these exclusive contacts in MDA::ER cells (the BS15/*TFF3* association). In contrast with what would have been expected from an artificial cellular model as compared to a more “natural” one constituted by MCF-7 cells, this indicates that the ER-mediated three-dimensional (re-)organization of the TFF cluster response to E2 is more important and more intricated in MDA::ER cells. These assays further showed that the ERBS located within the TFF promoter (comBS2) was directing more interactions than others in both cell lines. Hence, it may play crucial role in these spatial regulations of the TFF cluster.

Dynamic 3D organization of the TFF cluster. The dynamics of the loops linking ERBS to genes promoters were next evaluated by 3C-qPCR following E2 treatment of the cells. These data are summarized within **Figure 3**,

with circle areas being directly proportional to the fold changes in the relative frequency of interactions as compared to the initial situation (t0) (MCF-7 data are illustrated in **Figure S9**). In both cell types, a spatial reorganization was apparent as soon as 10 min following E2 addition with both increases and decreases in frequencies of interactions (**Figure 3**). Such variations in the three-dimensional organization of the TFF genomic region were also apparent in DNA-FISH experiments (**Figure S10**). Some of the reorganizations evidenced by 3C-qPCR exhibited a relatively dynamic or even cyclical nature, such as those highlighted in orange within Figure 3 subpanels. Interestingly, the dynamics of these spatial reorganizations differed between MDA::ER and MCF-7 cells. Hence, in MCF-7 cells, these variations were restrained to short-range interactions except for *TFF1/BS1*, *TFF1/BS18* and *TMPRSS3/BS8* contacts. The MDA::ER dynamic interactome in contrast highlighted important variations in long-range interactions between ERBSs and the promoters of regulated genes. This is notably illustrated by the interactions made by the ERBSs located in the 5' region of the cluster that climax at 50 min following the addition of E2. Furthermore, in MDA::ER cells, an apparent combination of local (BS2 and comBS1) and long-range interactions (comBS2 and even comBS3 located in the far 3' of the TFF genomic region) could be correlated with the E2-mediated regulations of *TMPRSS2* and *RIPK4*. In contrast to what happens in MCF-7, the regulations of core TFF genes would be more influenced by distal than by local ERBSs in MDA::ER cells. Indeed, variations of the interactions made between ERBSs located within the *TFF3* to *TMPRSS3* region appear less important than what was observed in the case of MCF-7 cells. In summary, these data indicate that there is no single major ERBS that organizes E2 responsiveness within this genomic region. Rather, coordination might originate from a combinatorial engagement of ERBSs located within the *TFF1* promoter in MCF-7 cells with two nearby ones (BS14 or BS16), and with those located in the distant 5' region of the cluster in MDA::ER cells.

Cell-specific E2 regulation of a given gene is driven by different ERBS. Although highly detailed, above 3C data do not precisely establish the exact contribution of each ERBS toward the specific regulation of the genes that they contact. In an attempt to obtain further experimental clues on this question, we used small triplex forming oligonucleotides (TFOs) to interfere with ER binding on its BS (**Figure 4A**). We characterized 11 TFOs (**Figure S11**) which were able *i*) to form DNA triplex *in vitro* (TFO anti-BS1 as exemplified in Figure 4B, otherwise see **Figure S12**); *ii*) to specifically bind to target sequences but not *PKNOX1* promoter used here as a control (**Figure 4C**; **Figure S12**) and *iii*) to significantly disrupt ER binding on the corresponding ERBS (**Figure 4D**, **Figure S13**). Subsequent RT-qPCR and 3C experiments showed that in the majority of the tested cases, decreasing ER binding on one ERBS did impact both the transcriptional status of the genes to which they were spatially close and the frequencies of these long-range interactions (**Figure 4E**, see also **Figures S14-S16**). These assays also indicated that the binding of ER on some distant sites might not be essential for the establishment of ERBS-promoter associations such as MCF-7 BS4/*UBASH3A*, BS6/*TMPRSS3* or MDA::ER BS10/*UBASH3A*. This also controls that the binding of a TFO to its target region does not in itself interfere nonspecifically with the 3C ligation step. Moreover, our control data also show that they do not hinder the digestion efficiency (not shown). In certain cases, reducing the frequencies of some of the characterized interactions had no statistical impact on the transcriptional state of the target gene. This was for instance true for the MCF-7 BS 10 with respect to *TMPRSS3*, BS14/*TFF2* and BS14/*TFF1*. Similar cases were detected in MDA::ER: BS1 for *TMPRSS2* and

TFF3, BS12 for *TMPRSS3* and *UBASH3A*. This indicates that these loops may have minor roles in mediating the E2-sensitiveness of these genes. Finally, these experiments evidenced that the abrogation of ER binding on some sites had side-effects on the transcriptional status of genes to which they are not linked by chromatin loops: the BS1 on *TFF3* levels, comBS1 on *TMPRSS2*, BS10 on *TFF2*, BS14 on *TMPRSS2* and *TFF3* in MCF-7 cells or BS10 on *RIPK4* and BS12 on *TMPRSS2*, *RIPK4* and *TFF3* amounts in MDA::ER. We were able to determine that all of these “collateral” effects might be due in part to the establishment of additional interactions between ERBS themselves (**Figure S17**). The use of TFOs therefore allowed us to confirm the functional relevance of the chromatin loops linking ERBS to E2-regulated promoters characterized in our assays. These results are summarized into the cytoscape layouts in **Figure 4F** and **4G** that represent respectively MCF-7 and MDA::ER functionalized interactomes of the TFF cluster in which the thickness of the edges is proportional to the functional importance of the interaction in terms of gene regulations. Calculations made to define a functional parameter are within **Supplementary Tables 2** and **3** and took into account the impact of the TFOs on gene expression but also on ER recruitment and frequencies of interactions. Although limited to the ERBS on which TFOs were able to disrupt ER recruitment, these interactomes illustrate, as discussed before, that the comERBS2 located in the close vicinity of the *TFF1* promoter plays a central regulating role in both cell types. They also clearly indicate that the MDA::ER ERBS1 plays a prominent role in *TFF1* gene expression.

CTCF and cohesin organizes the estrogenic responses of the TFF cluster. Following this functional characterization of MDA::ER and MCF-7 interactomes linking ERBS to gene promoters, we next aimed at deciphering the processes that may be at the origin of these cell-specific spatial (re-)organizations of the TFF chromatin domain. We first wondered whether the mobilization of CTCF and/or cohesin, proteic partners involved in the three-dimensional organization of chromatin (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008) on cell-specific ERBS would participate to these processes. We first mapped CTCF and RAD21 (a subunit of the cohesin complex) binding sites in MDA::ER through ChIP-chip analysis (**Figure 5A**). The comparison of MDA::ER CTCFBS and RAD21BS with those previously determined in MCF-7 cells (Schmidt et al., 2010) indicated that most CTCFBS were conserved between the two cell types unlike RAD21BS (**Figure S18**). This cell-specific RAD21 cistrome may represent a major source for the different E2-responses of the clustered TFF genes. On the other hand, the overlap between the BS for ER, CTCF and RAD21 was much more important in MCF-7 cells than in MDA::ER within the TFF cluster (**Figure 5B**). This observation was confirmed by Re-ChIP experiments (data not shown), and was also evident on all regions spotted within the array (**Figure S18**). There was no significant variations in CTCF and RAD21 binding on respective BSs following the addition of E2 (**Figure 5C**; **Figure S19**) that could explain the dynamic modulations of the TFF three-dimensional organization. These differing ER/CTCF/RAD21 BS overlaps between MCF-7 and MDA::ER were reflected in the proportions of established contacts between ERBS and genes promoters (**Figure 5D**). These data also indicated that the presence of CTCF and/or RAD21 at ERBS does not correlate with the number of interactions made with gene promoters in the context of the TFF cluster (Chi2 and Fisher tests p-values>0.4).

Despite these limited overlaps between ER binding sites and those of CTCF and RAD21, disrupting CTCF and RAD21 expression by siRNAs (controls in **Figure S20**) drastically reduced both basal and induced transcriptional

activity of E2-regulated genes in both cell lines (**Figure 5E**). This ultimately led to a strong decrease in their fold inductions by E2 except for *UBASH3A* in MCF-7 and *TFF3* and *RIPK4* in MDA::ER cells. CTCF and RAD21 are thus likely exerting an important function in the E2-responsiveness of the TFF cluster genes. Whether this was related to a role of these proteins in the spatial conformation of this genomic domain was next assayed in DNA-FISH experiments (**Figure 5F**). Unfortunately, the poor depletion in CTCF protein ($\approx 25\%$) obtained when following the particular cell culture procedure required for these experiments prevented any definitive conclusion to be made concerning the exact role of CTCF. However, these assays clearly demonstrated that RAD21 is essential for the estrogen-induced constraints exerted on the TFF three-dimensional conformation observed in control siRNA conditions. Accordingly, the silencing of RAD21 but also CTCF, although to a limited extent, diminished the frequency of interactions between ERBS and their target gene promoters as determined from 3C-qPCR experiments (**Figure 5G**). Altogether, these data indicate that CTCF and RAD21 organize the transcriptional responses of the genes included in the TFF cluster to E2 in both cell lines.

GATAs and FOXA1 have different roles in controlling the TFF cluster activity. We next further inquired the nature of the processes involved in the cell-specific organization of the TFF domain by questioning whether different pioneer or regulating factors might define the MCF-7 and MDA::ER cell-specific ER cistromes. We focused on FOXA1 and GATA factors because of the observed *i*) functional interplay between ER and GATA3 in MCF-7 cells (Eeckhoutte et al., 2007); *ii*) enrichment of FOXA1 and GATA binding motifs in MCF-7 ERBS and GATA2 motifs in MDA::ER ERBS; and *iii*) the absence of FOXA1 in the MDA-MB231 cell context (**Figure S21**; Eeckhoutte et al., 2006). We first complemented available MCF-7 ChIP-chip data for FOXA1 and GATA3 (Carroll et al., 2006; Hua et al., 2009) with ChIP-chip profiling of GATA3 recruitment in MDA::ER cells and that of GATA2 in both cell lines (**Figure 6A**). In contrast to what was observed with GATA3, ChIP-chip analysis of GATA2 in MCF-7 cells did not evidence any specific enrichment of this protein on ERBS (**Figure S21**). Next experiments therefore focused on GATA3 only in this particular cell line. GATA factors were found present on 7 or 5 ERBS in MCF-7 or MDA::ER cells, respectively (**Figure 6B**), a lower overlap than what was observed for FOXA1 and ER in MCF-7. This was not specific to the TFF cluster genomic region (**Figure S21**). Re-ChIP assays (data not shown) confirmed the overlaps of ER, GATAs and/or FOXA1 BSs in MCF-7 and MDA::ER cells with the exception of the MDA::ER ERBS 13 on which the recruitment of GATA2 and GATA3 might be mutually exclusive.

To obtain further details on the actions of these proteins, we reduced their expression using siRNAs (controls in **Figure S20**). Silencing of FOXA1 in MCF-7 or GATAs in both cell lines reduced ER binding on each tested sites, but the effect of the FOXA1 siRNA was clearly the most drastic (**Figure 6C**). As observed previously (Eeckhoutte et al., 2007), there was a small reduction in the amounts of mRNAs encoding ER in MCF-7 cells following the transfection of siRNAs directed against GATA3. However, this reduction was not translated at the protein level at the times used in our experiments (**Figure S20**). According to the pioneering functions of FOXA1 (Lupien et al., 2008; Serandour et al., 2011), the chromatin accessibility of the majority of the MCF-7 ERBS did reduce in the absence FOXA1, as monitored by FAIRE experiments (**Figure 6C**). This was not correlated to changes at the levels of H3K4 mono- and di-methylation (me1/me2) status, as shown by Lupien et al. (2008). Surprisingly, silencing GATA3 enhanced the chromatin accessibility of ERBS1,9 and comERBS1 in conjunction

with a significant gain in H3K4me2 but not H3K4me1. Finally, there was also a 40 to 50 % loss in H3K27ac on sites where ER binding was reduced following the reduction in FOXA1 and GATA3 proteins (Figure 6C). In MDA::ER cells, reducing GATA2 or GATA3 amounts did reproduce such a pattern of ER and H3K27ac loss occurring in conjunction with a gain of FAIRE and H3K4me2 enrichments (**Figure 6C**). But, in contrast to MCF-7 cells, chromatin modifications provoked by the silencing of GATA2 and GATA3 were accompanied by a significant gain of H3K4me1. These data altogether indicate that, in MCF-7 cells, FOXA1 and GATA3 have diverging -if not inverse- functions in regulating the chromatin status of ERBSs.

At the transcriptional level, except for *UBASH3A*, FOXA1 extinction reduced the fold induction of E2-sensitive genes included in the TFF cluster in MCF-7 (**Figure 6D**), as expected from the observed drastic reductions of ER binding. siRNAs directed against GATA3 also reduced responses to E2 of all genes but *UBASH3A*. However, contrasting with the FOXA1 siRNA effects, the reduced fold-inductions following the transfection of the GATA3 siRNA were mainly a consequence of an enhanced basal activity. In MDA::ER cells, diminishing GATA2 and GATA3 also provoked strong variations in the basal activity of E2-regulated genes, but these opposed to what was observed in MCF-7 cells. These strong inhibitory effects led to a diminished fold induction for *TFF3* only. These data indicated that GATA factors have important roles on the basal transcription state of E2-regulated genes, in both MCF-7 and MDA::ER cells, although with an inverse output. To consider this hypothesis, we performed ChIP and FAIRE assays as previously but included chromatin prepared from untreated cells (**Figure 6E**). Importantly, in both cell lines, the silencing of GATA2 and/or GATA3 enhanced the binding of ER on these GATA positive ERBS in the absence of E2 as compared to control siRNA, thereby reducing the fold-change in ER mobilization (**Figure 6E**). Correspondingly, the enhanced engagement of ER was correlated with a 2 to 3 log₂ unit-fold enhancement of FAIRE signals in the absence of E2. In contrast, FOXA1 silencing in MCF-7 cells diminished chromatin accessibility and ER binding in the absence of E2 and also reduced or even abrogated GATA3 mobilization on corresponding sites (data not shown). Together with the observation that the GATA3 siRNA did not impact FOXA1 engagement, these data demonstrate that GATA3 function is dependent upon a preliminary action of FOXA1 in MCF-7 cells. These data also show that GATA proteins restrict ER activity in MCF-7 and MDA::ER cells in the absence of E2 by limiting the extent of chromatin modifications that take place around the ERBS. We also demonstrated by 3C-qPCR (**Figure 6F**) that in the majority of the cases where silencing GATA factors had an effect on the spatial organization of the TFF cluster in the absence of E2, it reflected those observed in E2 treated cells transfected with the control siRNA. Exceptions were the comERBS2/*TFF3* and ERBS20/*TFF2* interactions in MCF-7 or ERBS2/*TFF2* and ERBS6/*RIPK4* in MDA::ER cells. By restraining the binding of ER in the absence of E2, GATA proteins seem therefore able to prevent the dynamic spatial reorganization of the TFF locus from taking place when the hormone signal is absent.

DISCUSSION

We used here a combination of different experimental procedures to generate a dynamic portrayal of the molecular processes allowing a hormone to regulate the transcriptional activity of clustered genes. Focusing on a large genomic region including the TFF cluster, we compared the nature of these processes in a naturally E2-responsive breast cancer cell line (MCF-7) and in a cellular system with a forced E2-sensitivity (MDA::ER), in order to interrogate whether these mechanisms could be triggered by ER on its own. Our data are in agreement with previous observations regarding the involvement of CTCF and RAD21 in the establishment of key connections between distant regions and regulated promoters of the TFF cluster and the key role of FOXA1 in organizing estrogen signaling (Schmidt et al., 2010; Zhang et al., 2010). But, in addition, they provide an unprecedented level of understanding of the mechanisms involved in the spatial and temporal regulation of this genomic region.

Although confronted to different chromatin contexts, ER was found to drive tight regulations of the TFF cluster in both cell types. Despite inducing a rather similar transcriptional output in MCF-7 and MDA::ER, the actions of ER were relayed by distinct genomic regions. Preparation of chromatin to ER binding involves the actions the pioneer factor FOXA1 in MCF-7 cells (Carroll et al., 2006). However, FOXA1 is not expressed in MDA::ER cells, suggesting that additional pioneers might be involved. Interestingly, the chromatin status of MDA::ER ERBSs in native MDA-MB231 cells indicated that a number of regions was already exhibiting marks for poised or active chromatin (data not shown). Furthermore, some of the interactions detected in 3C were detected in MDA-MB231 cells. These observations indicate that ER can bind opportunistically to opened chromatin regions, at least for some of the ERBS of the TFF loci. Examining the existence of specific collaborating/pioneering factor and their roles in the functions of the TFF cluster, we observed complex interrelations between ER, FOXA1 and GATA3 in MCF-7, imitated by an ER/GATA2/GATA3 network in MDA::ER. From our data and those of others (Lupien et al., 2008), we could propose models for the interrelations allowing FOXA1 and/or GATA factors to control ERBS functions (**Figure 7**). Importantly, we found that GATA proteins tightly control the functionalization and functions of the ERBS located within the TFF cluster. This observation and the proposed model have obviously to be validated genome-wide. However, one of its outcomes would be that, in the absence of E2, the FOXA1+GATA3 positive ERBSs have limited chromatin accessibility as compared to the FOXA1 only, which is what we observed in FAIRE experiments (Figure S6). In MDA::ER cells, another factor would have to replace FOXA1 roles on the initial chromatin functionalization brought by H3K4me1/me2. GATA mediated events would be essentially the same as those occurring in MCF-7 cells but with an opposed final outcome on ER-mediated regulations. This signifies that an uncharacterized factor is lacking in the MDA::ER system to allow enhanced transcriptional activities of unliganded ER when GATA are absent. This protein could be a partner of FOXA1 in MCF-7 cells that would be able to overcome the lack of H3K27ac for instance. Alternatively, GATA2 or GATA3 may have additional coactivation properties in MDA::ER. Interestingly, such a dual model for GATA roles in transcriptional regulation has emerged in a recent study integrating distant interactions with the cistromes of many proteins (Lan et al., 2012) and, notably, a co-recruitment of chromatin remodeling complexes such as SWI/SNF with GATA2.

4C and 3C assays allowed the description of 4 dimensional interactomes defining ERBSs-promoter interactions for genes which were estrogen-sensitive in the majority of the cases. Our MCF-7 interactome is in accordance with the ERBS centered one identified by ER ChIA-PET (Fullwood et al., 2009), but uncovered additional interactions of ERBS with promoters. The MCF-7 and MDA::ER interactomes exhibited different dynamics over time following treatment with E2, with an E2-mediated reorganization apparently less important in MCF-7 cells. It could be hypothesized that the chromatin three-dimensional structure of this region is already prepared for a response to E2 in MCF-7 cells, in contrast to the reconstituted E2-sensitive cellular model provided by MDA::ER cells. If true, this implies that in MDA::ER cells, ER on its own is able to provoke important three-dimensional remodeling to finely tune the transcription of target genes. This is in accordance to the decondensation of large scale chromatin structure previously evidenced (Nye et al., 2002).

Some of the detected interactions engaged promoters of genes which were not regulated by E2 in our RT-qPCR or transcriptomic data, such as *TMPRSS2*, *RIPK4* or *TFF3* in MCF-7 cells. As evidenced by global run-on assays (GRO-seq; Hah et al., 2011) it can however be possible that these genes would exhibit rapid transcriptional responses to estrogen (10 to 40 min following E2 treatment). Hence, the sole action of ER tethered to proximal or distant regions of this gene appears not sufficient to trigger processes required for its activation in MCF-7 cells. This is in line with a general problematic that emerged from genome-wide studies, which evidenced that cell-specificity and/or number of interactions made between the BSs of a transcription factor and genes promoters are not systematically reflected at the transcriptional level (Eeckoute et al., 2009). The use of TFOs gave important and novel clues on this particular aspect of deciphering ER actions and gave further indications on how such a high number of binding sites for a TF can be organized in space and time in order to regulate the transcription of its target genes. Indeed, perhaps one of the most novel concepts that emerge from our data is that ER-bound enhancers apparently exhibit a relative functional redundancy. Indeed, disruption of ER binding to certain ERBSs did not have any detectable impact on the transcriptional state of its distant target gene. As evidenced by our 4C experiments, a single promoter can establish contacts with several ERBSs, and the resulting buildout might be insensitive to a particular ERBS inactivation, providing robustness to the regulatory system. Alternatively, the mobilization of ER on distinct master regulatory regions appears sufficient in other cases to provoke the transcriptional response of the gene. Comparing the enrichments of main vs secondary ERBSs in particular histone modifications, ER, Pol II, CTCF or RAD21 proteins, or even their relative frequencies of interaction with their target promoters, did not evidence a particular segregating characteristic. We hypothesize that the differential recruitment of yet uncharacterized cofactors might account for these differences.

As for ER, the BSs for many transcription factors have been shown to be grouped around responsive genes in clusters of enhancers (Gotea et al., 2010). Hence, TFOs or other methods that interfere with the mobilization of TF are promising approaches to assess for the generality of enhancers redundancy

EXPERIMENTAL PROCEDURES

Detailed protocols are described in the Extended experimental procedures.

Chemicals, Enzymes, Antibodies and Recombinant Plasmids. All chemicals, restriction or modification enzymes were obtained from Sigma, Roche or New England Biolabs. Antibodies (**Table S4**) were purchased from Abcam, Diagenode, Millipore, NeoMarkers or SantaCruz, or gift from Dr. JM Peters (anti-Smc3 subunit of the cohesin complex). Enhancers were PCR-amplified (primers in **Table S5**) from genomic DNA prepared from corresponding cells and cloned into the CpG-free pCpGLCMV/EF1 vector (Klug and Rehli 2006) using the In-Fusion cloning system (Clontech) and sequenced.

TFOs. We used a home-written python algorithm following the rules defined in Vekhoff et al. (2008) to design putative TFOs. Triplex formation was monitored *in vitro* by incubating increasing amounts of TFOs with DNA duplexes for 16h at 37°C in a buffer containing 10 mM MgCl₂, 100 mM NaCl, 50 mM Tris-HCl (pH 7.4), 10% glycerol and 0.5 mg/ml tRNA. Complexes were separated by native electrophoresis and visualized by methylene blue staining.

Cell culture and transfections. MCF-7, MDA-MB231 and MDA-MB231 cells stably expressing ER α (MDA::ER) (Reid et al., 2003) were maintained in DMEM (Gibco) containing 5% fetal calf serum (FCS, BioWest) and antibiotics (Roche) at 37°C under 5% CO₂. MDA::ER media was supplemented with 0.8 mg/ml hygromycin (Calbiochem). For experiments requiring treatment with E2, cells were cultivated for 2 days in DMEM/2% charcoal-stripped FCS (csFCS; BioWest) prior to the addition of E2 (10⁻⁸M final concentration). One μ mol siRNAs (**Table S6**) or 10 μ mol TFOs were transfected for 36h using oligofectamin as recommended by the manufacturer (Invitrogen) in FCS and antibiotics-free OptiMEM (Gibco).

DNA-Fluorescent in situ Hybridization (FISH). DNA Probes were produced from BACs (Invitrogen) through random priming (Bioprime array CGH genomic labeling system, Invitrogen). Permeabilized PFA fixed cells were subjected to denaturation through sequential heating in 70% formamid/30% 2X SSC and 50% formamid/50% 2X SSC. Hybridization with labeled denatured probes was performed overnight at 42°C in the presence of Cot-I (Invitrogen) and salmon sperm DNA (Sigma). Measurements of nucleus area and distance between the centroid of each probe were performed under ImageJ (<http://rsbweb.nih.gov/ij/>). Images from up to 150 nuclei were analyzed in each experiment. Distances were determined in 2D, since pilot experiments did not evidence any qualitative difference between 2D and 3D-FISH experiments (data not shown). To calculate 3D volumes, we first segmented automatically the 3-dimensional hybridization signals for each of the color channel using the triangle algorithm (Zack et al., 1977) implemented in ImageJ. After a cleaning step consisting of the successive application of an opening and closing filter, the 3D-volumes of the structures resulting from the union of the two segmentation masks were measured and expressed as voxels.

Reverse-Transcription. Total RNAs from 10⁷ cells were purified using TrizolTM reagent (Life Technologies, Inc.) according to the manufacturer's instructions. Two μ g of RNA served as template for M-MLV reverse transcriptase (Invitrogen) and Pd(N)6 random hexamers (Amersham Pharmacia Biosciences).

Chromosome conformation capture (3C) and circular 3C (4C). Methods were adapted from Gondor et al.

(2008) and Hagege et al. (2007), and used the *DpnII* 4-base cutter as an enzyme of choice. Detailed protocol and experimental workflows are illustrated in the Extended Experimental Procedures and in Figure **S22**.

DNA-IP, FAIRE and ChIP assays. Me-DIP protocol, adapted from Mohn et al., 2009, used 4 µg of fragmented DNA and 2 µg of specific antibody, as described in Serandour et al., 2011. Formaldehyde-assisted isolation of regulatory elements (FAIRE) methodology was conducted as previously (Serandour et al., 2011). ChIP experiments were conducted with minor modifications from previous protocols (Metivier et al., 2003). ChIP material was amplified using the WGA whole genome amplification kit (Sigma) for arrays hybridization. All subsequent steps were performed by NimbleGen services (Reykjavic, Island). Triplex capture assays were performed on transfected MCF-7 or MDA::ER cells which were subsequently cross-linked by 2% formaldehyde and lysed by sonication using a BioRuptor apparatus (Diagenode). Sonicated chromatin was then incubated for 4h with streptavidin-coated magnetic beads (Dynal) that were blocked with 10 µg/ml BSA and 10 µg/ml yeast tRNA for 1h. Captured DNA was eluted by two rounds of elution in 0.1% SDS, purified following digestion with proteinase K and RNaseA and analyzed by qPCR.

Quantitative PCR (qPCR). All qPCRs were performed on BioRad CFX96 machines using BioRad iQ SYBR Green supermix. Primers for RT-PCR were designed under QuantPrime (<http://www.quantprime.de/>; Arvidsson et al., 2008). Oligonucleotides for all other type of samples were designed under Primer3 (<http://frodo.wi.mit.edu/primer3/>). Sequences for all primers (Sigma) are in **Tables S7 to S10**.

Bioinformatics and statistics. ChIP-chip normalization and peak calling steps were performed using the MA2C algorithm (Song et al., 2007) on raw data obtained from two custom NimbleGen arrays hybridized with DNA prepared in two independent triplicate experiments. MCF-7 data for genome-wide ER, Pol II, FOXA1 and GATA3 ChIP-chip were taken from Carroll et al., 2006 and Hua et al., 2009. CTCF and cohesin ChIP-seq fastq data from the GSE25710 were re-analyzed as described in Extended experimental procedures. Genomic annotations and motifs searches were performed under the cistrome platform (<http://cistrome.dfci.harvard.edu/ap>; Blankenberg et al., 2007). Networks of long-distance chromatin interactions were generated under Cytoscape (Shannon *et al.*, 2003). Heatmaps were all generated under Mev (Saeed et al., 2003), with values that were declared as significant from the control by Wilcoxon or Student t-tests (depending upon the number of values). To facilitate the reading of the heatmaps, the value of an experimental sample was replaced by the control one when no significant difference was calculated.

ACCESSION NUMBERS

The MDA::ER and MCF-7 data generated during this study have been deposited in NCBI's Gene Expression Omnibus (Barrett et al., 2009) and are accessible through GEO Series accession numbers GSE32132 ChIP-chip data) and GSE23850 (transcriptomes).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures with associated Supplementary references, 22 Figures and 10 Supplementary Tables.

ACKNOWLEDGEMENTS

This work was supported by the CNRS, the University of Rennes I, and benefited from grants of the ARC, the Ligue Contre le Cancer (Equipe Labellisée Ligue 2009), Région Bretagne (CREATE 4793) and the the Agence Nationale pour la Recherche (ANR-09-BLAN-0268-01). J.Q. is a recipient of a joint fellowship by the CNRS and the Region Bretagne. A.S was supported by a fellowship from the French Ministère de l'enseignement supérieur et de la recherche, and S.C. was funded by a CNRS post-doctoral fellowship. We greatly acknowledge Dr. Denis Habauzit for all his advices and help on the design and analysis of TFOs.

AUTHOR CONTRIBUTIONS

All biological experiments performed by JQ, CLP, SC, MB, GP, AS, FP, SP, VL and RM were conceived and designed by RM, VL, JQ, EW, SC, JE and GS. SA programmed all local informatic tools to analyze whole-genome data, and SH those for the DNA-FISH experiments. RM and JQ wrote the manuscript with the help of all other authors.

REFERENCES

- Arvidsson, S., Kwasniewski, M., Riano-Pachon, D. M., and Mueller-Roeber, B. (2008). QuantPrime--a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* 9, 465.
- Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M., Marshall, K. A., et al. (2009). NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37, D885-890.
- Blankenberg, D., Taylor, J., Schenck, I., He, J., Zhang, Y., Ghent, M., Veeraraghavan, N., Albert, I., Miller, W., Makova, K. D., et al. (2007). A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. *Genome Res* 17, 960-964.
- Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoutte, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F., et al. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* 38, 1289-1297.
- Carroll, J.S., and Brown, M. (2006). Estrogen receptor target gene: an evolving concept. *Mol. Endocrinol.* 20, 1707-1714.
- Carvajal, J.J., Keith, A., and Rigby, P.W. (2008). Global transcriptional regulation of the locus encoding the skeletal muscle determination genes *Mrf4* and *Myf5*. *Genes Dev.* 22, 265-76.
- Chambeyron, S., and Bickmore, W.A. (2004). Does looping and clustering in the nucleus regulate gene expression? *Curr. Opin. Cell Biol.* 16, 256-262.
- Chinery, R., Williamson, J., and Poulsom, R. (1996). The gene encoding human intestinal trefoil factor (TFF3) is located on chromosome 21q22.3 clustered with other members of the trefoil peptide family. *Genomics.* 32, 281-284.
- Eeckhoutte, J., Keeton, E.K., Lupien, M., Krum, S.A., Carroll, J.S., and Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res.* 67, 6477-6483.
- Eeckhoutte, J., Métivier, R., and Salbert, G. (2009). Defining specificity of transcription factor regulatory activities. *J Cell Sci.* 122, 4027-4034.

- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., et al. (2009). An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 462, 58-64.
- Giresi, P. G., Kim, J., McDaniel, R. M., Iyer, V. R., and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* 17, 877-885.
- Gondor, A., Rougier, C., and Ohlsson, R. (2008). High-resolution circular chromosome conformation capture assay. *Nat Protoc* 3, 303-313.
- Gotea, V., Visel, A., Westlund, J.M., Nobrega, M.A., Pennacchio, L.A., and Ovcharenko, I. (2010). Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res* 20: 565-577.
- Hah, N., Danko, C.G., Core, L., Waterfall, J.J., Siepel, A., Lis, J.T., and Kraus, W.L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* 145, 622-634.
- Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2, 1722-1733.
- Hogan, G. J., Lee, C. K., and Lieb, J. D. (2006). Cell cycle-specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet* 2, e158.
- Hsu, P. Y., Hsu, H. K., Singer, G. A., Yan, P. S., Rodriguez, B. A., Liu, J. C., Weng, Y. I., Deatherage, D. E., Chen, Z., Pereira, J. S., et al. (2010). Estrogen-mediated epigenetic repression of large chromosomal regions through DNA looping. *Genome Res* 20, 733-744.
- Hua, S., Kittler, R., and White, K.P. (2009). Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell* 137, 1259-1271.
- Hurst, L. D., Pal, C., and Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5, 299-310.
- Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll JS. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* 43, 27-33.
- Lan, X., Witt, H., Katsumura, K., Ye, Z., Wang, Q., Bresnick, E.H., Farnham, P.J., and Jin, V.X.. (2012). Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* [Epub ahead of print] doi: 10.1093/nar/gks501.
- Li, Q., Barkess, G., and Qian, H. Chromatin looping and the probability of transcription. *Trends Genet.* 22, 197-202.
- Lupien, M., Eeckhoute, J., Meyern C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell.* 132, 958-970.
- Magnani, L., Ballantyne, E.B., Zhang, X., and Lupien, M. (2011). PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS Genet.* 7:e1002368.
- McEwan, I. J. (2009). Nuclear receptors: one big family. *Methods Mol Biol* 505, 3-18.
- Metivier, R., Gallais, R., Tiffocche, C., Le Peron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature*

452, 45-50.

- Metivier, R., Penot, G., Hubner, M. R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor- α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115, 751-763.
- Mohn, F., Weber, M., Schubeler, D., and Roloff, T. C. (2009). Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol* 507, 55-64.
- Motallebipour, M., Ameer, A., Reddy Bysani, M.S., Patra, K., Wallerman, O., Mangion, J., Barker, M.A., McKernan, K.J., Komorowski, J., and Wadelius, C. Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq. *Genome Biol.* 10, R129.
- Nicol, J. W., Helt, G. A., Blanchard, S. G., Jr., Raja, A., and Loraine, A. E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25, 2730-2731.
- Nye, A. C., Rajendran, R. R., Stenoien, D. L., Mancini, M. A., Katzenellenbogen, B. S., and Belmont, A. S. (2002). Alteration of large-scale chromatin structure by estrogen receptor. *Mol Cell Biol* 22, 3437-3449.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. et al. (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422-433.
- Perkins, T. J., Hallett, M., and Glass, L. (2004). Inferring models of gene expression dynamics. *J Theor Biol* 230, 289-299.
- Reid, G., Hubner, M. R., Metivier, R., Brand, H., Denger, S., Manu, D., Beaudouin, J., Ellenberg, J., and Gannon, F. (2003). Cyclic, proteasome-mediated turnover of unliganded and liganded ER α on responsive promoters is an integral feature of estrogen signaling. *Mol Cell* 11, 695-707.
- Rubio, E.D., Reiss, D.J., Welcsh, P.L., Distech, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008) CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* 105, 8309-8314.
- Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M. et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374-378.
- Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* 20, 578-588.
- Serandour, A. A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R., et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.*
- Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A., and Brown, M. (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* 103, 843-852.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13, 2498-25504.
- Song, J. S., Johnson, W. E., Zhu, X., Zhang, X., Li, W., Manrai, A. K., Liu, J. S., Chen, R., and Liu, X. S. (2007). Model-based analysis of two-color arrays (MA2C). *Genome Biol* 8, R178.

- Tan, S.K., Lin, Z.H., Chang, C.W., Varang, V., Chng, K.R., Pan, Y.F., Yong, E.L., Sung, W.K, and Cheung, E. (2011). AP-2 γ regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *EMBO J.* 30, 2569-2581.
- Tang, Y., Huang, Y., Shen, W., Liu, G., Wang, Z., Tang, X.B., Feng, D.X., Liu, D.P., and Liang, C. C. (2008). Cluster specific regulation pattern of upstream regulatory elements in human alpha- and beta-globin gene clusters. *Exp. Cell Res.* 314, 115-122.
- van Driel, R., Fransz, P.F., and Verschure, P.J. (2003). The eukaryotic genome: a system regulated at different hierarchical levels. *J. Cell Sci.* 116, 4067-4075.
- Vekhoff, P., Ceccaldi, A., Polverari, D., Pylouster, J., Pisano, C., and Arimondo, P.B. (2008). Triplex formation on DNA targets: how to choose the oligonucleotide. *Biochemistry* 47, 12277-12289.
- Welboren, W.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated? *Endocr. Relat. Cancer.* 16, 1073-1089.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T. et al. (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796-801.
- Zack, G.W., Rogers W.E., and Latt, S.A. (1977). Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* 25, 741-753.
- Zaret KS, and Carroll JS. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* 25, 2227-2241.
- Zhang, Y., Liang, J., Li, Y., Xuan, C., Wang, F., Wang, D., Shi, L., Zhang, D., and Shang, Y. (2010). CCCTC-binding factor acts upstream of FOXA1 and demarcates the genomic response to estrogen. *J. Biol. Chem.* 285, 28604-28613.

FIGURE LEGENDS

Figure 1. Dynamic coordinated regulations of E2-sensitive genes and global remodeling of the TFF locus

(A) Integrated genome browser (IGB; Nicol et al., 2009) illustration of the TFF genomic region. Locations of the BAC clones used to generate green and red fluorescent probes, chr21 coordinates and RefSeq genes are shown.

(B) Estrogenic regulations of the genes included within the TFF locus, as determined from expression microarray experiments (GSE23850). Red boxes indicate activated genes.

(C) Heatmap representation of RT-qPCR results obtained on RNAs prepared from MCF-7 and MDA::ER cells treated for the indicated times with 10^{-8} M E2. Results shown are the log2 of the fold inductions of the genes expression levels by E2, and were obtained in two independent triplicate experiments.

(D) Kinetic ChIP performed on MCF-7 or MDA::ER cells treated with E2 following the presence of activated RNA Polymerase II (P-Pol) on the promoter regions of indicated genes. Data shown are fold enrichment towards control ChIP and control sequence.

(E) Representative pictures of a DNA-FISH experiment using a mix of all generated probes in cells treated with E2 or ethanol (EtOH) for 50 min. S1, S2 and S3 are the three TFF loci present in MDA-MB231 cells. The two pictures with higher magnification illustrate the compaction of this entire genomic loci upon E2 treatment. Quantitative analysis of the 3D volume of the loci (416 measures) is indicated at the bottom of the panel with indicated Fisher t-test p-value.

(F) Quantitative analysis of the distribution of distances separating paired FISH probes. Values are represented within boxplots (top) or within quantile-quantile (Q-Q) plots representation of conditionally ranked measured distances (bottom). The normal distribution expected from non-varying distances is illustrated by the straight black line in each Q-Q plots. Non-parametric Fisher test p-values are indicated when determined as significant, as well as the number of measurements made in at least three independent experiments.

Figure 2. Cell-specific three-dimensional organization of the TFF cluster

(A) IGB visualization of MCF-7, MDA-MB231 and MDA::ER interactomes linking *DpnII* fragments corresponding to ERBS to the promoter regions of indicated genes, as detected by 4C-qPCR following a 50 min treatment of the cells by 10^{-8} M E2. Shown are RefSeq genes coordinates along chr21, as well as the MA2C scores of an ER ChIP-chip analysis performed using chromatin of MDA::ER cells treated for 50 min with E2. MCF-7 MAT scores were obtained from published genome-wide ER ChIP-chip dataset (Carroll et al., 2006). For the sake of figure clarity, only the highest 5% of the MA2C or MAT signals are shown. Black and brown boxes delineate cell-specific ER binding sites (ERBS) whilst common ERBS (comERBS) are shown in green. 4C data are represented as lines linking one ERBS to its target promoters. MCF-7 interactions are in red, MDA::ER ones in grey and those shared between MDA::ER and MDA-MB231 in blue.

(B) Venn diagrams depicting the overlapping interactions characterized in this study and to those identified in the published ER ChiA-PET dataset (Fullwood et al., 2009) restricted to loops involving the gene promoters that served here as 4C baits.

(C) Stacked histograms illustrating the overlap of the 4C-detected interactions for each tested promoters in the different cell lines.

(D and E) Cytoscape (Shannon *et al.*, 2003) circular layouts of the networks of interactions that link E2-regulated genes to ERBS in MCF-7 (D) and in MDA::ER (E) cells. The sizes of the nodes are directly related to the number of interactions they direct.

Figure 3. Dynamic three-dimensional reorganization of the TFF cluster upon E2 treatment

Summary of one 3C experiment representative from two, performed on chromatin sampled from MCF-7 (A) or MDA::ER cells (B) treated from 0 to 80 minutes with 10^{-8} M E2. As indicated in the framed legend, the size of the bubble that corresponds to one interaction is proportional to the fold changes in frequencies of interaction as compared to the basal (t0) situation. The location of the genes promoters that served as anchors is illustrated on the top of each subpanel and the ER ChIP-chip profiles on the left. Distance scale is accurate (2 Mb between ticks) but had to be broken in some instances for sake of figure size and clarity. Bubbles highlighted in orange are those commented in the main text and those in yellow correspond to interactions made by the ERBS located within the *TFF1* promoter (comERBS2).

Figure 4. Functionalization of MCF-7 and MDA::ER interactomes using TFOs

(A) We designed triplex forming oligonucleotides (TFOs) to interfere with ER binding on single ERBS to disrupt and thus to identify the roles of ER on this specific genomic region for the regulation of E2-sensitive genes.

(B) Formation of DNA triplex analyzed by gel-shift. Increasing amounts of TFO (25 to 1500 pmol) were added to 25 pmoles of target DNA duplexes and incubated for 16h at 37°. A control was made by incubating duplexes with an unspecific TFO (Ctrl TFO) at the highest concentration. Complexes were separated by electrophoresis and stained with methylene blue.

(C) MCF-7 cells were transfected for 36 h with 10 µmol of TFO or biotinylated (Biot-) TFO directed against the ERBS1, subjected to cross-linking and sonicated chromatin was then incubated with streptavidin-coated beads. Amounts of captured DNA were analyzed by qPCR. Values shown are mean ± SD of two independent duplicates, and are expressed as % of captured DNA relative to input DNA normalized to the amounts of recovered negative control region (*Rplp0* promoter).

(D) Anti-ER Chip-qPCR performed on MCF-7 cells transfected as previously and treated for 4h with 10^{-8} M E2. Results are mean values ± SD of three independent triplicates expressed as relative enrichment towards the *PKNOX1* promoter used as negative control for ER binding.

(E) Anti-ER ChIP-qPCR, 3C-qPCR and RT-qPCR experiments performed on MCF-7 and MDA::ER cells transfected as in D. 4C boxes indicate identified interactions. Experimental values were normalized to those obtained in untransfected cells and expressed as log2. ChIP and RT data originate from three independent duplicate or triplicate experiments, whilst those from 3C are of one representative experiment from two.

(F and G) The experimental data reported in E were used to refine MCF-7 (F) and MDA::ER (G) interactomes by adjusting the strength of the interconnections of both MCF-7 and MDA::ER networks to their relative functional relevance which is reflected by the thickness of the edges. The size of the nodes of these cytoscape layouts is proportional to the number of interactions that they direct.

Figure 5. CTCF and cohesin roles in organizing the E2-reponsiveness of the TFF cluster

(A) MCF-7 (Carroll *et al.*, 2006; Schmidt *et al.*, 2010) and MDA::ER anti-ER, CTCF or RAD21 ChIP-Seq or ChIP-chip

signals visualized under IGB as in figure 2. MDA::ER ChIP-chip data were obtained from cells treated with E2 for 50 min.

(B) Overlap of ER, CTCF and cohesin BS on the TFF genomic region in MCF-7 and MDA::ER cells.

(C) The effect of E2 on the recruitment of CTCF and RAD21 was evaluated by ChIP (bottom). ChIP results are log2 mean values of three independent triplicates expressed as relative enrichment towards the *PKNOX1* promoter.

(D) Repartition of CTCF and RAD21 BS within the identified interactions between ERBS and genes promoters.

(E) Heatmaps summarizing RT-qPCR results obtained in at least 3 independent triplicate experiments following the transfection of siRNAs directed against CTCF or RAD21 expression and a 4h treatment of the cells with 10^{-8} M E2 or EtOH as vehicle control. A siRNA targeting the Luciferase (siLuc) mRNA was used as a control. Expression levels were normalized to the *PKNOX1* internal control and reported to those calculated for the siLuc EtOH condition. The fold inductions of these genes expression by E2 are also shown on the right side of the panel as relative to those measured in cells transfected by the siLuc. Results originate from at least 4 independent triplicate experiments.

(F) Quantitative DNA-FISH analysis of the distribution of distances separating paired fluorescent probes in transfected MDA::ER cells. Results are illustrated as in Figure 1.

(G) 3C-qPCR experiments following the impact of the reduction in CTCF and RAD21 contents on the interactions between indicated ERBS and genes promoters. Results shown originate from one representative experiment from two.

Figure 6. FOXA1 and GATA factors have different roles in the estrogenic regulation of the TFF cluster

(A) MCF-7 FOXA1 and GATA3 cistromes (data from Carroll et al., 2005; Hua et al., 2009) and MDA::ER GATA1, GATA2 and GATA3 ones determined by ChIP-chip assays on 50' E2-treated cells, visualized as in Figure 2.

(B) Overlaps between FOXA1 and/or GATA proteins in MCF-7 and MDA::ER cells.

(C) FAIRE and ChIP-qPCR experiments following the siRNA-mediated reduction of FOXA1 and GATA proteins. These assays were performed on MCF-7 and MDA::ER cells treated for 50' with 10^{-8} M E2. Data originating from anti-ER ChIP are expressed as fold enrichment relative to an ER negative sequence. FAIRE data were normalized against a control region exhibiting a closed chromatin conformation. Anti-modified histones ChIP were normalized to control ChIP (beads alone) and then to those obtained in an anti-unmodified H3 ChIP. Both FAIRE and anti-modified H3 ChIP were further compared to the values obtained in the siLuc control. All experiments were performed at least twice in triplicates.

(D to F) Heatmaps summarizing RT-qPCR (D), ChIP and FAIRE (E) or 3C (F) data obtained in experiments performed following the transfection of MCF-7 or MDA::ER cells with indicated siRNAs and treatment with 10^{-8} M E2 or EtOH for 50 min. RT-qPCR results were obtained in at least 3 independent triplicate experiments. Normalizations were done as described in Figure 5. The relative mobilization of ER, relative chromatin status of these genomic regions and 3C data were calculated as previously and are shown as compared to the values obtained in samples originating from cells transfected with the control siLuc siRNA.

Figure 7. Suggested interplay between GATA proteins and FOXA1 that prepares chromatin for ER actions

This scheme summarizes the hypothetical mechanisms by which FOXA1 and GATA proteins might orientate the cell-specific regulations of the TFF cluster. In MCF-7 cells, binding of FOXA1 on nucleosomes that could be marked by H3K4 monomethylation and low level of dimethylation would be essential for the successive recruitment of a methyl-transferase that enhances K4 dimethylation. GATA3 would then intervene to limit the extent of these modifications. In parallel, the establishment of a basal H3K27 acetylation state would functionalize the ERBS for the mobilization of unliganded ER. Hormone binding would then stabilize the frequencies of ER-ERBS interactions and provoke regulations of target gene through modifications of chromatin nature and spatial organization. In the absence of GATA3, the action of the FOXA1 dependent methyl-transferase would be sustained, leading to an enhanced chromatin accessibility facilitating the mobilization of unliganded ER. This would allow ER to regulate target gene basal activities, but to a lower extent than in control conditions in the presence of E2. This could be associated with the uncompleted functionalization of the ERBS apparent from the lower levels of H3K27 acetylation.

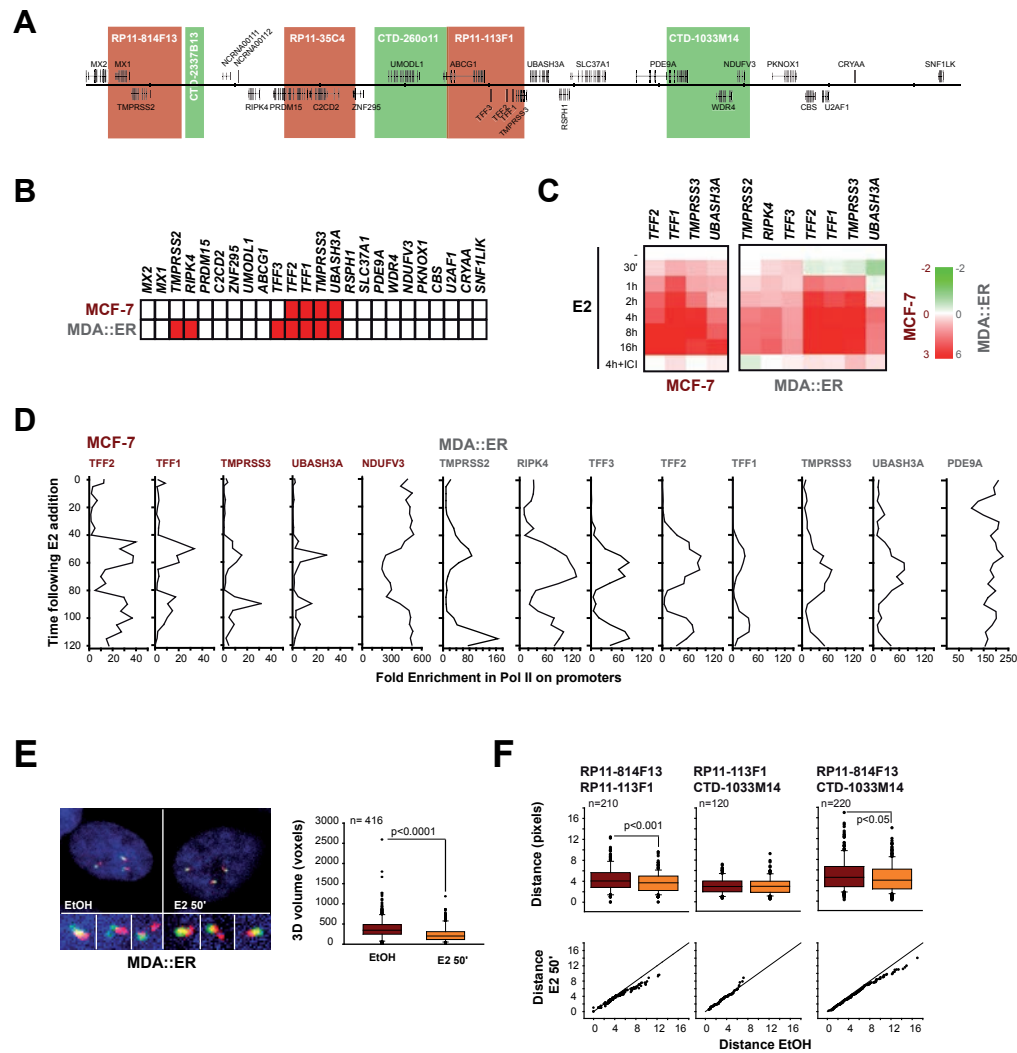
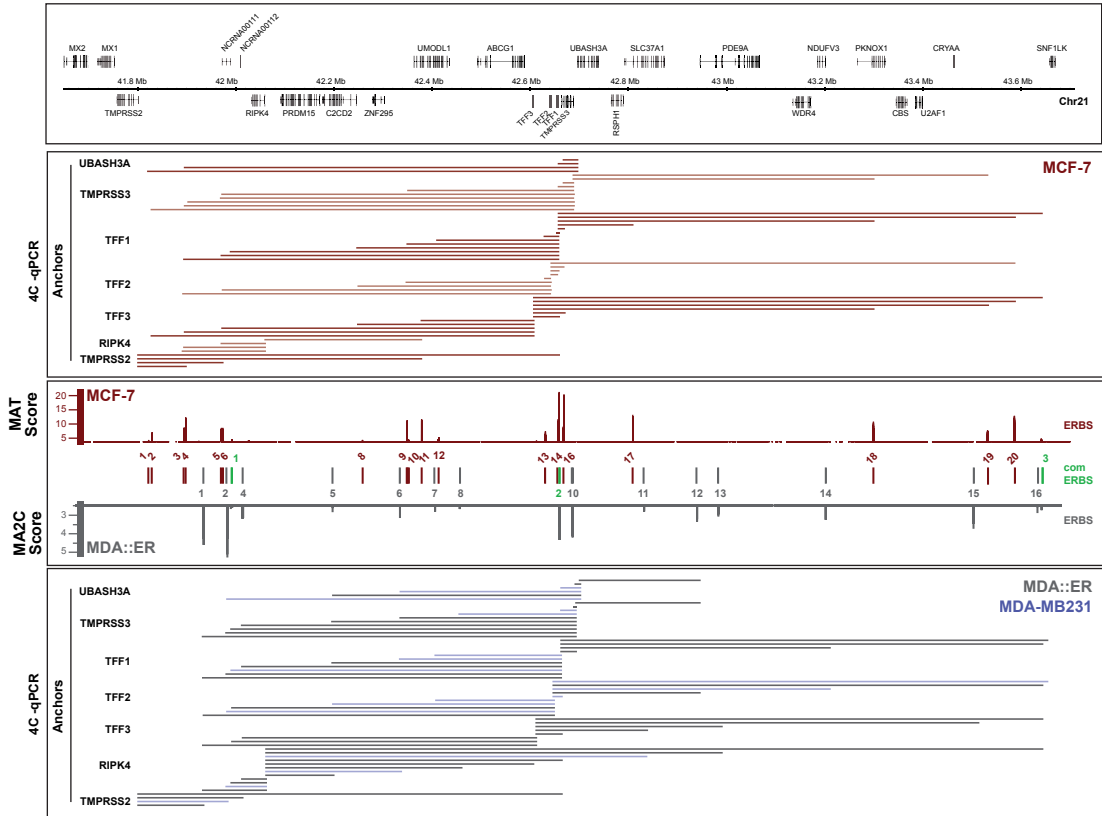
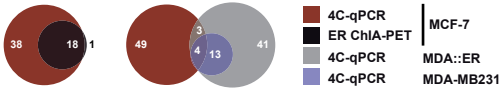


Figure 1

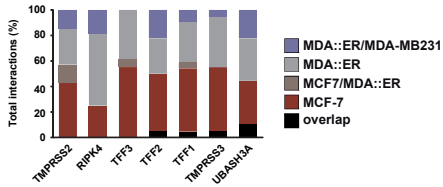
A



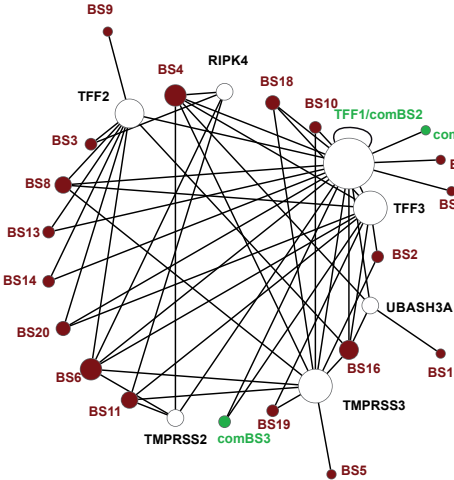
B



C



D



E

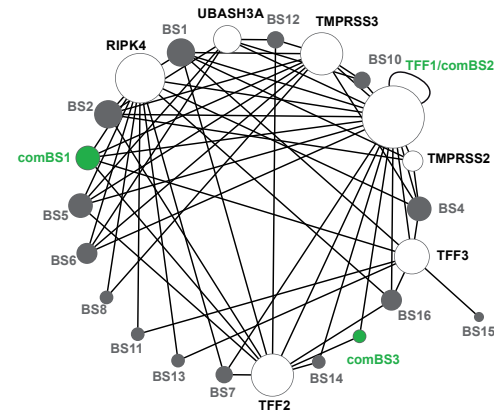
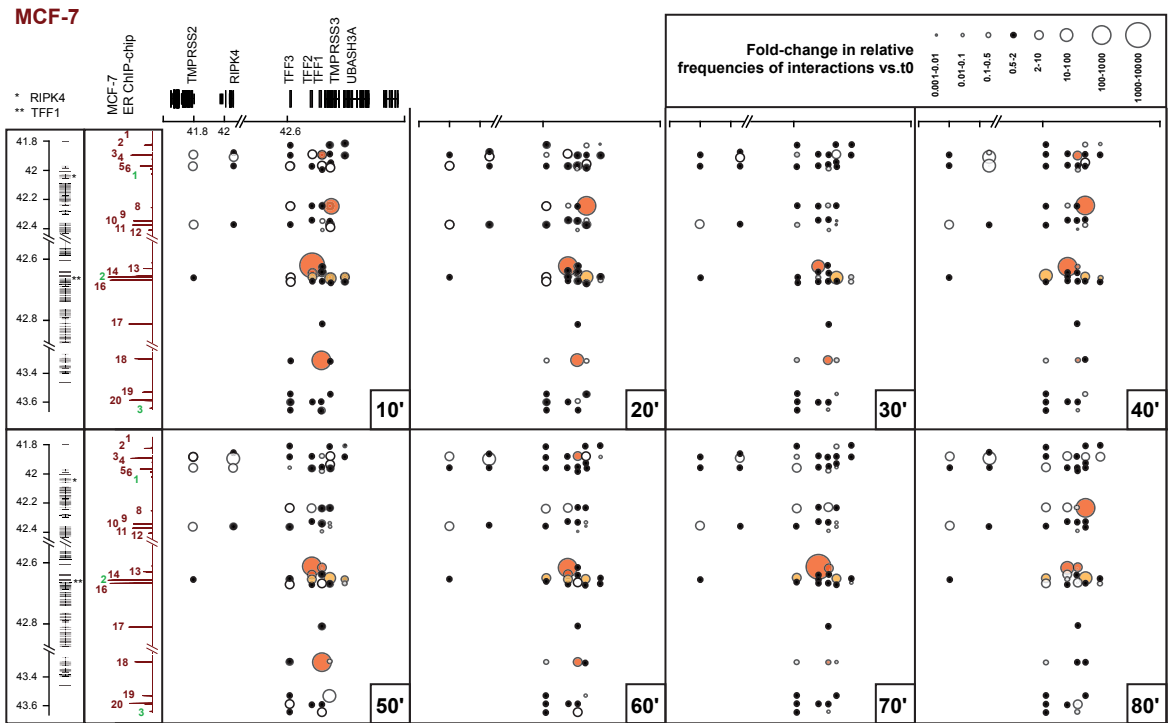


Figure 2

A



B

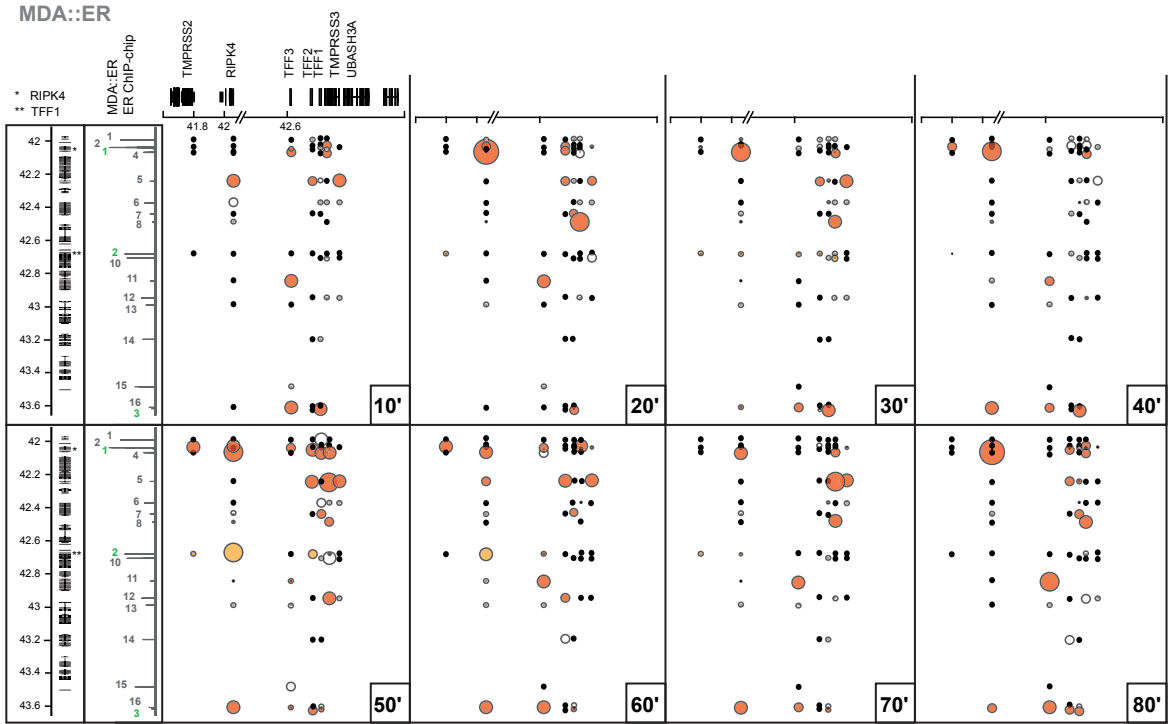


Figure 3

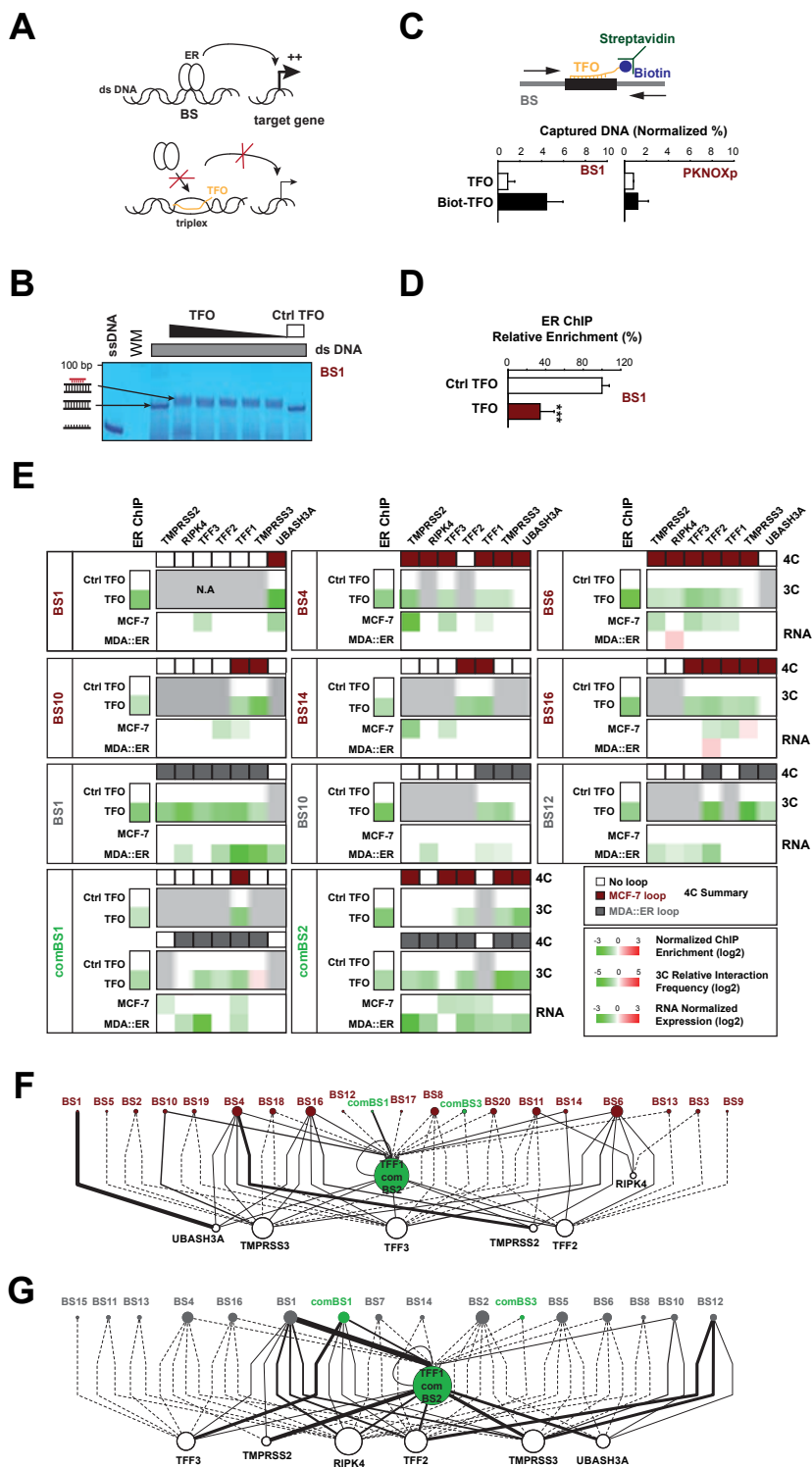
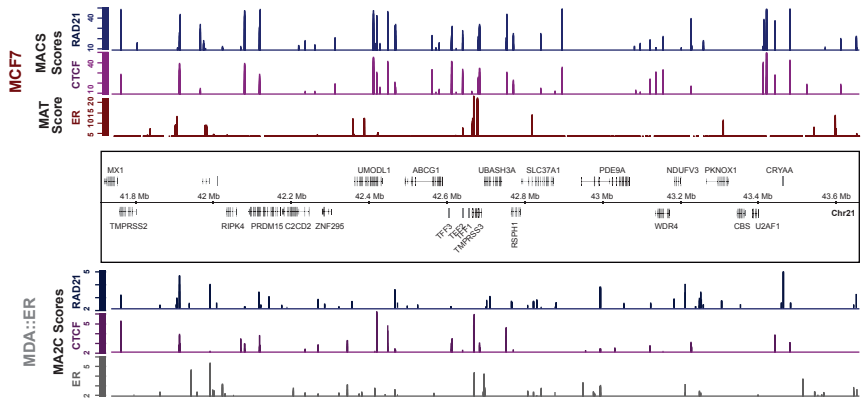
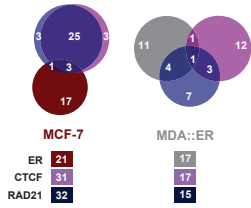


Figure 4

A



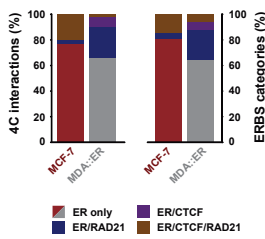
B



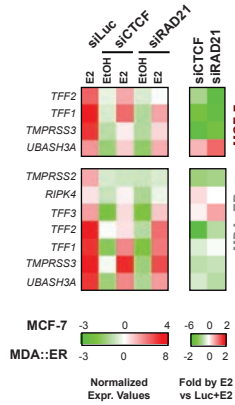
C



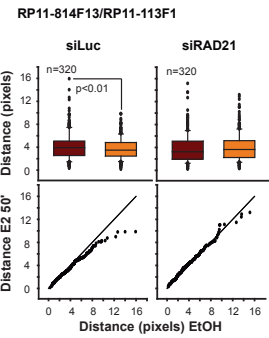
D



E



F



G

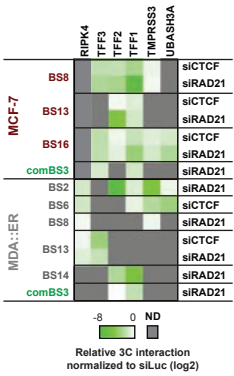


Figure 5

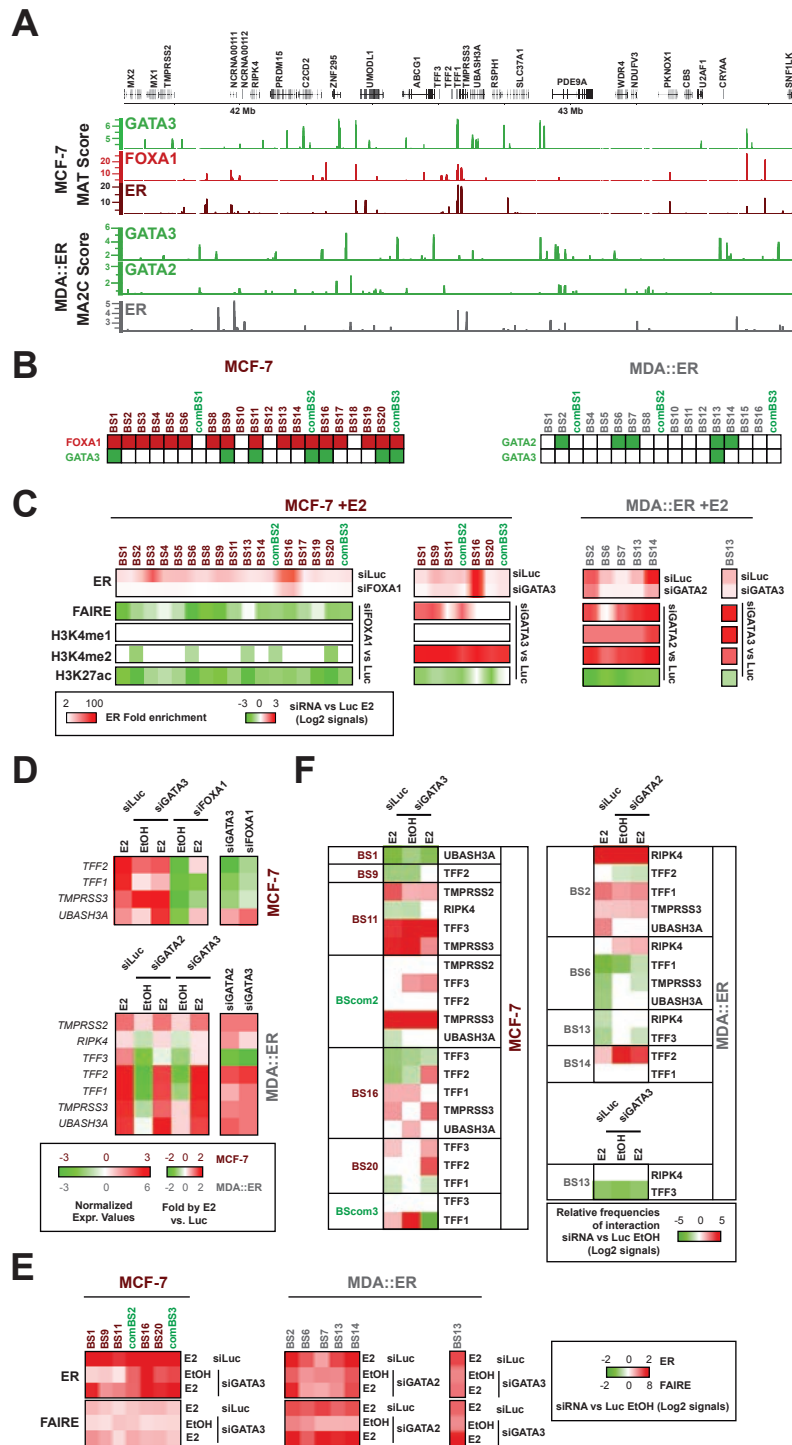
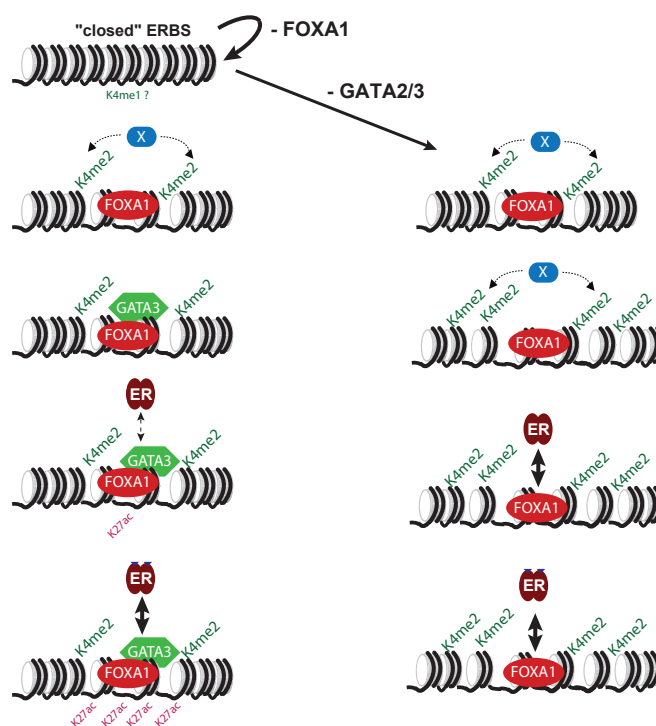


Figure 6



SUPPLEMENTARY MATERIAL FOR:**CELL-SPECIFIC AND SPATIO-TEMPORAL CONTROLS OF THE ESTROGEN-RESPONSIVE TREFOIL FACTOR (TFF) LOCUS ACTIVITY**

Justine Quintin¹, Christine Le Péron¹, Gaëlle Palierne¹, Maud Bizot¹, Stéphanie Cunha^{1,*}, Aurélien A. Sérandour^{1,**}, Stéphane Avner¹, Frédéric Percevault¹, Sébastien Huet¹, Erwan Watrin³, Jérôme Eeckhoutte^{1,4}, Vincent Legagneux², Gilles Salbert¹, and Raphaël Métivier¹

¹ Equipe SP@RTE, UMR CNRS 6290. Equipe labellisée Ligue contre le Cancer Université de Rennes I. Campus de Beaulieu. 35042 RENNES CEDEX. FRANCE. ² Equipe EGD, UMR CNRS 6290. Université de Rennes I. Faculté de Médecine. 2 avenue du Professeur Léon Bernard. 35043 RENNES CEDEX. FRANCE. ³ Equipe CC, UMR CNRS 6290. Université de Rennes I. Faculté de Médecine. 2 avenue du Professeur Léon Bernard. 35043 RENNES CEDEX. FRANCE. ⁴ INSERM U1011, Université Lille-Nord de France, Faculté de Médecine de Lille-Pôle Recherche. 59045 Lille. FRANCE.

Contact:

Raphaël Métivier, Equipe SP@RTE, UMR CNRS 6290. Batiment 13. Université de Rennes 1. Campus de Beaulieu. 35042 Rennes Cedex. Tel: +33 (0)2 2323 5142 Fax: +33 (0)2 2323 6794; E-mail: raphael.metivier@univ-rennes1.fr

Additional footnotes

* Present Address: Huntsman Cancer Institute. Salt Lake City, Utah. USA.

** Present Address: CRI, Cancer Research UK, Cambridge. UK.

EXTENDED EXPERIMENTAL PROCEDURES

Cell culture. MCF-7, MDA-MB231 and MDA-MB231 cells stably expressing ER α (MDA::ER) (Reid et al., 2003) were maintained in DMEM (Sigma) supplemented with 5% fetal calf serum (FCS, Sigma) and antibiotics (Roche) at 37°C under 5% CO₂. MDA::ER cells were grown in DMEM/5% antibiotics and hygromycin (Calbiochem, 0.8 mg/ml). For experiments that required cells to be treated with E2, cells were cultivated for 3 days in 9cm \emptyset dishes in DMEM/0.5% charcoal-stripped FCS (csFCS).

DNA-FISH. Probes were produced by direct labeling of BACs clones (Invitrogen) through random priming (Bioprime array CGH genomic labeling system, Invitrogen) using fluorochrome conjugated nucleotides (dUTP-alexa fluor 488 from Invitrogen or dUTP-cyanine 3 from Perkin Elmer). Before use, probes were denatured by heating 5min at 80°C and then 30min at 37°C. Cells were grown for two days on glass slides in DMEM without phenol red containing 2.5% csFCS. After addition of 10⁻⁸M E2 or ethanol (vehicle), slides were washed with PBS containing Ca²⁺ and Mg²⁺, and then fixed in 2% paraformaldehyde (PFA) for 10 min at 4°C. PFA fixed cells were permeabilized in 0.5% Triton X-100 and equilibrated in 1X SSC for 5 min. Slides were incubated one hour with 20ug/mL RNaseA in 1X SCC at 37°C and then sequentially washed 3 times with PBS, incubated in 2% PFA for 10 min at room temperature (RT), in HCl 100mM for 10 min RT and then in 0.5% Triton for 10 min RT, with 3 washings with PBS between each step. Slides were then subjected to denaturation through sequential heatings at 73°C in 70% formamide/ 30% 2X SSC for 7 min and then 3 min in 50% formamide / 50% 2X SSC. Hybridization with 600ng of labeled denatured DNA probes was then performed overnight at 42°C in hybridization buffer (per 800 μ L: 200 μ L 25% Dextran sulfate; 100 μ L 20X SSC; 500 μ L deionised formamid) containing 150 μ g of Cot-I (Invitrogen) and 150 μ g of Salmon Sperm DNA. Slides were rinsed three times in 2X SSC, in 50% formamide/50% 2X SSC 20 min at 42°C and three times in 2X SSC again. Nuclei were stained with DAPI in 2X SSC for 5 min and then slides were mounted in ProLong Gold antifade reagent (Invitrogen) with a 22*40 coverslip.

Microscopy and Image analysis. Stacks were obtained with a 63x oil immersion objective of a DMRXA (LEICA) microscope or a Zeiss apotome (63x objective). Measurements of nuclear area and distance between the centroid of each probe were performed under the free software Image J (<http://rsbweb.nih.gov/ij/>). Entire stacks were taken for all selected nuclei (non-mitotic and containing the expected 3 pairs of hybridization signals), and the three channels (red, green and blue) were isolated using the «DeInterleave» plugin. Pictures in z containing maximum red or green signal intensities were selected for all channels, merged and then thresholded to eliminate background from specific signals for distance measures. We used the DAPI (blue channel) pictures to consistently determine the nucleus area, calculated following the determination of a threshold fluorescence value corresponding to an entry transition into the nucleus. This value was manually determined as the inflection point of a profile plotting DAPI signal measured in a 10 pixel large longitudinal window crossing the nucleus against pixel distance. Images from up to 100 nuclei were analyzed in each experiment. Significant variations between experimental conditions were tested by a Fisher *t-test* comparison for unpaired data, with a significance threshold set for p-values ≤ 0.05 . To analyse the kinetic FISH experiment, we had to develop a custom Matlab (MathWorks, Natick, Massachusetts) image processing routine in order to quantitatively analyze the high number of images. For that, we used the maximum intensity projections of the

3-dimensional stacks acquired by fluorescence microscopy. The analysis steps were performed automatically to avoid potential bias associated to manual intervention. The nuclei were segmented on the DAPI channel using the Otsu approach (Otsu, 1979) and a watershed algorithm (Meyer, 1994) was applied to separate the touching nuclei. For detecting the fluorescence spots on the images corresponding to each FISH probe, we used the algorithm developed by Sbalzarini and Koumoutsakos (2005). A first filtering step was performed to remove the very dim spots and those located outside the nuclei. The intensities of the remaining spots were estimated and subtracted to the local background. We only kept the spots whose intensity was exceeding the threshold TI, which was calculated as follows:

$$TI = \langle I \rangle + n \cdot \sigma I,$$

with $\langle I \rangle$ and σI , the mean and the standard deviation of the spot intensities, respectively, and n a user-defined integer. On a subset of images, we compared the spots detected by the automatic analysis and those selected by manual inspection. By setting n to 5, we optimized the matching between the automatic and manual selections. The intensity spots corresponding to the two FISH probes were paired based on a nearest neighbor criterion. FISH pair distances exceeding 20 pixels were not considered. We noted that our approach was robust towards the choice of the integer n as similar distance distributions were obtained for $3 < n < 7$.

To calculate 3D volumes, we first segmented automatically the 3-dimensional hybridization signals for each of the color channel using the triangle algorithm (Zack et al., 1977) implemented in ImageJ. After a cleaning step consisting of the successive application of an opening and closing filter, the 3D-volumes of the structures resulting from the union of the two segmentation masks were measured and expressed as voxels.

siRNA transfections. 2.5×10^6 MDA::ER cells were plated in 9cm dishes in DMEM/5%FCS for 16h and then grown for 24h in DMEM/2.5% csFCS. Media was then replaced with 4 ml of FCS free Opti-MEM (Sigma) and 1 μ mol siRNAs (Table S6) were then transfected using oligofectamine as recommended by the manufacturer (Invitrogen). Following 6h of incubation, the media was completed with 125 μ l of csFCS, and E2 stimulation (10^{-8} M) was done 36h later.

Plasmid transfections. To assay for their putative enhancer activities, selected regulatory regions (approx. 2 kb surrounding MDA::ER or MCF-7 ERBS) were amplified by PCR from genomic DNA (primers in Table S5) prepared from corresponding cells, and cloned into the CpG-free luciferase expression vector pCpGLCMV/EF1 (Klug and Rehli, 2006). For luciferase reporter assays, 20,000 cells were seeded in 24-well plates. Twenty four hours after seeding, cells were transfected in triplicates with 250 ng of reporter genes and 50 ng of Renilla vector (pRL-Null, Promega) in diluted JetPEI reagent (Polyplus Transfection). Twenty-four hours following transfection cells were treated with 10^{-8} M E2 during 4 h before determination of Firefly and Renilla luciferase activities (Dual-Luciferase Reporter Assay System, Promega).

RT-PCR. Total RNAs from 10^7 cells were purified using the TrizolTM reagent (Life Technologies, Inc.) according to the manufacturer's instructions. Two μ g of RNA were used as template for M-MLV reverse transcriptase (Invitrogen) and Pd(N)6 random hexamers (Amersham Pharmacia Biosciences). All qPCR reactions used 1/10th

of the reverse transcription reaction. Independent triplicate or duplicate experiments were conducted three times, and all values were normalized to *Rplp0* or *PKNOX1* mRNA.

Microarray and mRNA profiling data analyses

RNA for microarray analysis were isolated from 20^7 MCF7 or MDA::ER cells treated or not with E2 using the RNeasy Plus Mini Kit (Qiagen) with homogenization through QIAshredders (Qiagen). Integrity and purity of the RNAs were controlled on an Agilent Bioanalyzer using the RNA 6000 Nano Assay (Agilent). Ten μ g of selected samples exhibiting a RIN >9.8 were then subjected to cDNA synthesis using the Superscript Double-Stranded cDNA Synthesis Kit (Invitrogen) and a mix of 50 pmol random hexamers and 50 pmol of Oligo-dT. cDNAs were then treated for 10 min at 37°C with 5 μ g RNaseA (Invitrogen), purified through a phenol:chloroform:isoamyl alcohol extraction on Phase Lock light gels (Fisher Scientific) and then precipitated. Following a quality control on agarose/BET gels, all subsequent steps (labeling of cDNA, hybridization and scanning of the arrays) were performed at the NimbleGen service facilities (Rejkyavic, Island). For each experimental condition, three arrays (NimbleGen Homo sapiens 385K Array) were hybridized with independently prepared pools of cDNAs synthesized from experimental triplicates (independent experimental and biological triplicates). Quantile normalization of the data through the RMA algorithm and all primary analyses were performed using the ArrayStar software suite (DNASTar, Inc.). Data were filtered according to two criteria: i) expression values greater than the first quartile in all samples of at least one triplicate; and ii) triplicate standard deviations lower than the third quartile in all triplicates. Experimental groups were compared by analysis of variances (t-test) and p-values were adjusted by the Benjamini and Hochberg (BH) method. Genes were considered as differentially expressed between two experimental conditions when their adjusted p-value was lower than 0.05 and their fold change greater than 1.8.

Me-DIP assays. Protocol was adapted from (Mohn et al., 2009). Genomic DNA was extracted from cultured cells using the DNeasy Blood and Tissue kit (Qiagen) with a supplementary step of RNase digestion. Before carrying out MeDIP, genomic DNA was sonicated to produce random fragments ranging from 200 to 500 bp in size. We used 4 μ g of fragmented DNA for a standard MeDIP assay, which were heated for 10 min at 95°C and then immunoprecipitated overnight at 4°C with 2 μ g of antibody directed against 5-methylcytidine (MAb-5MECYT, Diagenode) in a final volume of 500 μ L IP buffer (10 mM sodium phosphate pH 7.0, 140 mM NaCl, 0.05% Triton X-100). Immune complexes were recovered using 40 μ L of BSA coated Protein A Sepharose beads for 2 h at 4 °C. Following 3 washes with 1 mL of IP buffer, complexes were incubated with proteinase K overnight at 55 °C in a digestion buffer (50mM Tris HCl pH 8.0, 10mM EDTA, 0.5% SDS). Methylated DNA was then recovered by phenol-chloroform extraction followed by an ethanol precipitation.

Cross-linked chromatin preparations. Cells were washed twice with PBS, and cross-linked during 10 min at room temperature using 1.5% formaldehyde (ACROS) diluted in PBS. Following a subsequent incubation with 0.125 M glycine for 2 min, the cells were collected in 1 ml collection buffer [100 mM Tris-HCl (pH 9.4) and 100 mM DTT]. Cells were then washed in 1 ml PBS, lysed for 15 min at room temperature in 300 μ L of lysis buffer [10 mM EDTA, 50 mM Tris-HCl (pH 8.0), 1% SDS, 0.5% Empigen BB (Sigma)], and chromatin sonicated during 14

min using a BioRuptor apparatus (Diagenode), with 30 sec on/off cycles. Chromatin was then cleared through a 10 min centrifugation at 10,000 x g.

Chromatin immunoprecipitation (ChIP) assays. A tenth of the chromatin samples (30 µl) of the supernatants were used as inputs, and the remainder diluted 5-fold in IP buffer [2 mM EDTA, 100 mM NaCl, 20 mM Tris-HCl (pH 8.1), and 0.5% Triton X-100]. 1/4th of this fraction was subjected to immunoprecipitations overnight after a 3 hr preclearing at 4°C with 10 µg yeast tRNA and 80 µl of a 50% protein A-Sepharose bead (Amersham Pharmacia Biosciences) slurry. Complexes were recovered after 3 hr incubation at 4°C with 5 µg yeast tRNA and 40 µl of protein A-Sepharose. Precipitates were then serially washed, using 300 µl of Washing Buffers (WB) I [2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 0.1% SDS, 1% Triton X-100, 150 mM NaCl], WB II [2 mM EDTA, 20 mM Tris-HCl (pH 8.1), 0.1% SDS, 1% Triton X-100, 500 mM NaCl], WB III [1 mM EDTA, 10 mM Tris-HCl (pH 8.1), 1% NP-40, 1% Deoxycholate, 0.25 M LiCl] and then twice with 1 mM EDTA, 10 mM Tris-HCl (pH 8.1). Precipitated complexes were removed from the beads through two sequential incubations with 50 µl of 1% SDS, 0.1 M NaHCO₃. Cross-linking was reversed by an overnight incubation at 65°C. DNA was purified on NucleoSpin™ columns (Macherey-Nagel) using 50 µl NTB buffer. Subsequent Q-PCR analysis used 1 µl of input material and 3 µl of ChIP samples. In sequential-ChIP assays, DNA/protein complexes were extracted from the beads twice by adding 50 µl of 10 mM DTT and incubation for 20 min at 37°C. Supernatants were then diluted 10 times in IP buffer, and the IP procedure was performed again.

ChIP-on-chip assays. The ChIP procedure conducted on chromatin prepared from 15.10⁶ cells was similar to that described above with the following modifications for final steps. Crosslinking was reversed by an overnight incubation at 65°C with 10 µg of Proteinase K (Sigma). Following a subsequent incubation of the samples with 2.5 µg RNase (Sigma) for 1h at 37°C, the DNA was then purified on NucleoSpin™ columns (Macherey-Nagel) using NTB buffer and eluted in 40 µl of elution buffer. Efficiency of the ChIP assay was then evaluated using qPCR positive and negative controls. Experimental input and ChIP triplicates were then pooled by precipitation, resuspended in 25 µl H₂O, and divided in two for amplification using the WGA whole genome amplification kit (Sigma). Following a quality control step, the amplified material was pooled and sent to NimbleGen services (Reykjavic, Island) for hybridization on custom 385K arrays

Formaldehyde-assisted isolation of regulatory elements (FAIRE). A tenth of chromatin samples (30 µl) were used as inputs, and the remainder subjected to three consecutive phenol/chloroform/isoamyl-ethanol (25:24:1) extractions. Inputs and extracted samples were then incubated overnight at 65°C to reverse formaldehyde cross-linking with 10 µg of proteinase K. DNA was then purified on NucleoSpin™ columns (Macherey-Nagel) using NTB buffer, and quantified on a Nanodrop apparatus (Thermo Scientific). Twenty ng of inputs and purified DNA were used in Q-PCRs.

Chromosome conformation capture (3C) and circular 3C (4C). The walkthrough for the 3C and 4C experiments is detailed within **Figure S22**. Following a 5 min centrifugation at 2,000 x g, aliquots of 2.10⁶ cross-linked cells were washed with 200 µl of 1X *DpnII* buffer, and then lyzed overnight at 37°C in 200 µl of 1X *DpnII* buffer containing 0.3 % SDS with shaking. Following 2 passages through a syringe needle, 400 µl of 1X *DpnII* buffer were sequentially added in 4 times, and SDS was sequestered with 67 µl of 20 % Triton X-100 at 37°C for 1h. 50

µl of the reaction mixture were then kept as input fraction for digestion efficiency controls. 550 µl of the chromatin preparation were then digested overnight with 400 U *DpnII* at 37°C with shaking in a total volume of 500 µl of digestion buffer containing protease inhibitors (Roche). An additional step of 6h digestion with 150 U *DpnII* was then performed. 50 µl of the digested chromatin was then kept for digestion efficiency controls, whilst the remaining was kept at 4°C during this step. To control the digestion efficiency, both input and digested aliquots were incubated with 9.5% SDS for 20 min at 65°C. Cell fragments were then eliminated by centrifugation at 12,000 x g for 5 min. 117 µl of TE buffer were then added together with 5 µl of 10 µg/µl RNase A (Sigma) and the mixture was incubated at 37°C for 30 min before the addition of 8 µl of 5M NaCl and 10 µl of 10 µg/µl of Proteinase K (Sigma). Cross- linking was then reversed overnight at 65°C, and DNA purified on Macherey-Nagel columns. qPCR were then performed on input and digested fractions to calculate the digestion efficiency as follows:

$$E\% = [1.9^{(Ct_{input} - Ct_{sample})^{test\ region}} / 1.9^{(Ct_{input} - Ct_{sample})^{control\ region}}] * df * 100,$$

where the dilution factor (df) was =0.98360 [(50/610)/(50/600)], and the control region a region that contains no *DpnII* fragment. If this % was > 85%, the remaining digested chromatin was subjected to a final lysis step by addition of 108 µl of 10% SDS and incubation at 65°C with shaking for 20 min. Two µg of digested chromatin, as evaluated from the amounts of DNA recovered in the digested fraction using a NanoDrop were then incubated for 1h at 37°C with 40 µl of 20 % Triton X-100 in a total volume of 694 µl of Ligation buffer (10 mM Tris-HCl pH=7.8; 0.1 µg/µl BSA and protease inhibitors). Ligations for 3C or 4C experiments were then performed at 16 °C with gentle agitation for 4h or 5 days respectively. For 3C, 10 µl of T4 DNA ligase were added in the reaction mixture together with 80 µl of its buffer, 8 µl of 10 µg/µl BSA and 8 µl of 100 mM ATP (total volume of 800 µl/DNA concentration=2.5 ng/µl). For the 4C samples, requiring a more diluted concentration of DNA, 906 µl of H₂O were first added and then 160 µl of 10X T4 DNA ligase buffer, 16 µl of 10 µg/µl BSA, 16 µl of 100 mM ATP and 15 µl of T4 DNA ligase (total volume of approx. 1800 ml). The ligation mix was replenished at days 2 and 4 with ATP (20 µl of 100 mM ATP). Following these 5 days of ligation, 4C samples were further incubated for 1.5h with 1 ul of T4 DNA ligase in order to ensure that the ligase fills any nicks in the circularized 3C products. The cross-linking of either 3C or 4C DNA products was then reversed overnight at 65°C following the addition of EDTA (1 mM final), NaCl (0.2 M final) and 10 or 20 µl of 10µg/µl proteinase K. Samples were then subjected to three successive phenol/chloroform/isoamylalcohol (25:24:1) extractions, diluted 4 times in water and precipitated at -20°C for 2h by 2 volumes of isopropanol. Following centrifugation at 13, 000 x g, samples were then washed 3 times with 1 volume of 75% EtOH and resuspended in 50 µl of TE. 3C samples were then processed for analysis. In contrast, the circularized 3C products (4C) generated by the 5 day ligation were then purified from linear DNA by a combined digestion with 7 µl of exonuclease I and 2 µl of exonuclease III (New England BioLabs) in a total volume of 100 µl of 1X exonuclease I buffer. Circular DNA was then purified on Macherey-Nagel columns following a 25 min heating step at 85°C to inactivate the enzymes. Elution step was modified, by incubating the DNA bound on the columns with 50 µl of Tris-HCl (pH=7.8) containing 20 ng of yRNA for 2h. 4C libraries were then amplified on a thermocycler using a high-fidelity Taq polymerase from Invitrogen as described within Figure S30. In parallel of samples subjected to the whole procedure, additional

aliquots of cells or of chromatin were processed to generate additional controls: minus cross-linking (entire procedure on cells not incubated with formaldehyde), minus ligase (entire procedure but with no ligase in the mix). These samples served to control the specificity of the ligation (opportunistic ligation by background proximity: minus cross-linking) and of the PCR (minus ligase). The relative frequencies of interactions detected by 3C were calculated as follows:

$$\text{Freq} = [1.9^{(Ct_{\text{control region}} - Ct_{\text{test}})} / \min(E\%_{\text{test5'}}; E\%_{\text{test3'}})] / [1.9^{(Ct_{\text{control region}} - Ct_{\text{pos3Cctl}})} / \min(E\%_{\text{control}}; E\%_{\text{tpos3Cctl3'}})],$$

where the control region amplification (same as for the digestion controls) served to normalize over the inter-samples variations of DNA amounts, the pos3Cctl a region that was always ligating under any condition (ligation of two adjacent regions). All values were normalized for their availability for ligation, by taking into account the minimal (min) digestion efficiency measured for either extremities of the ligated product. As negative controls, we tested the minus ligase samples and assessed for two negative control interactions which were 1- the ligation of each fragments of interest with a fragment located within a *GAPDH* exon and 2- a negative region taken from our 4C screenings. We did not normalize the values obtained on test regions over those obtained in these negative controls as they were generally not producing any amplification signal. However, whenever this had to occur, the interaction detected with the test region was not taken into account. The 4C interactions were considered as being only qualitative, due to the amplification step present in the procedure. Calculations made to determine interaction regions were the same as for the 3C, except that in this case, the amounts of DNA were sufficient to allow further normalization over negative controls. In this case, the $Ct_{\text{control region}}$ used was the lower one (maximum interaction) obtained in the whole set of tested ligation-produced fragments. We considered values which were at least 2-fold higher than negative control regions as significant.

Bioinformatics. ChIP-chip signals normalization and peak calling steps were performed using the MA2C algorithm (Song et al., 2007) on raw data obtained from two arrays hybridized with DNA prepared in two independent experiments. MA2C parameters were: robust normalization ($C=1$) and peak calling for a minimum of 4 probes (maximum gap set at 400 bp) in a sliding window of 300 bp half-width. All binding sites determined in our ChIP-chip experiments were confirmed by independent ChIP-qPCR assays, except the CTCFBS identified in the TFF1 promoter which was found to be antibody- and experiment-dependent. This site was therefore not included in the statistical examinations of our data. Analysis of published Affymetrix tiling arrays data (ER, FOXA1 and GATA3 ChIP-chips performed in MCF-7 cells) were analyzed under MAT (Johnson et al., 2006). Genomic annotation of the MDA::ER and MCF-7 binding sites for all studied proteins distribution towards specific gene features was performed using the CEAS algorithm present within the cistrome web-platform (<http://cistrome.dfci.harvard.edu/ap>; (Blankenberg et al., 2007)). ChIP-seq data for CTCF and RAD21 and corresponding input were downloaded in fastq format from the arrayexpress database (<http://www.ebi.ac.uk/arrayexpress>; E-TABM-828), and aligned onto indexed chromosomes of genome hg18 using bowtie-0.12.7 (Langmead et al., 2009) with parameters -p 7, --best, -m 1, -n 1, --sam and -l 28. The .sam files were then converted to .bam files, using samtools-0.1.12a (Li et al., 2009) view, sort and index tools. The .bam files were themselves converted to .wig files using macs-1.3.7.1 (Zhang et al., 2008) with parameters --format=BAM --gsize 5130456 --tsize 28 --fold 10 --WIG.

TFOs. We used a home-written python algorithm following the rules defined in Vekhoff et al. (2008) to design putative TFOs targeting 15-30 bp long oligopyrimidine-oligopurine tracts included within ERBS, with one possible divergent base from a strict polyA/G sequence.

***In vitro* characterization of TFOs.** Triplex formation was monitored *in vitro* by incubating increasing amounts of TFOs with DNA duplexes for 16h at 37°C in a buffer containing 10 mM MgCl₂, 100 mM NaCl, 50 mM Tris-HCl (pH 7.4), 10% glycerol and 0.5 mg/ml tRNA. Complexes were separated by electrophoresis on 15% (19:1) polyacrylamide gel containing 10 mM MgCl₂ and 50 mM Tris-HCl (pH 7.4) and visualized by methylene blue staining. Triplex capture assays were performed on transfected MCF-7 or MDA::ER cells which were subsequently cross-linked by 2% formaldehyde and lysed by sonication as previously described. Sonicated chromatin was then incubated for 4h with streptavidin-coated magnetic beads (Dynal) that were blocked with 10 µg/ml BSA and 10 µg/ml yeast tRNA for 1h. Captured DNA was eluted by two rounds of elution in 0.1% SDS, purified following digestion with proteinase K and RNaseA and analyzed by qPCR.

Quantitative-PCRs (qPCR). All qPCRs used 1 µM of specific oligonucleotides (Sigma; sequences in Tables S7 and S8) and were performed on BioRad MyiQ and BioRad CFX96 machines using BioRad iQ SYBR Green supermix with 50 rounds of amplification followed by determination of melting curves. Primers for RT-PCR were designed using the QuantPrime design tool (<http://www.quantprime.de/>; (Arvidsson et al., 2008)). Oligonucleotides for all other type of samples were designed under Primer3 (<http://frodo.wi.mit.edu/primer3/>). Values were normalized in three steps: to inputs (Δ Ct), then to control ChIP samples (beads alone or anti-H3 ChIP; $\Delta\Delta$ Ct) and then to $\Delta\Delta$ Ct values obtained on control DNA regions.

SUPPLEMENTARY REFERENCES

- Johnson, W.E., Li ,W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. (2006). Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci U S A.* 103, 12457-12462
- Klug, M. and Rehli, M. (2006). Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics*, 1, 127-130.
- Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25, 2078-2079.
- Meyer, F. (1994). Topographic distance and watershed lines. *Signal Processing.* 38, 113-125.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Sys., Man., Cyber.* 9, 62–66.
- Sbalzarini I.F., and Koumoutsakos, P. (2005), Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* 151, 182-195.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

Table S1. Characterized interactions by 4C.

MCF-7	chr21	41822965	42697413	ERBS1_UBASH3A	0 +	41822965	42697413	0	2	1158,401	0,874048
	chr21	41830187	42608421	ERBS2_TFF3	0 +	41830187	42608421	0	2	422,573	0,777661
	chr21	41830187	42689569	ERBS2_TMPRSS3	0 +	41830187	42689569	0	2	422,986	0,858396
	chr21	41894208	42062049	ERBS3_RIPK4	0 +	41894208	42062049	0	2	1505,854	0,166987
	chr21	41894208	42643773	ERBS3_TFF2	0 +	41894208	42643773	0	2	1505,1891	0,747674
	chr21	41802859	41899656	ERBS4_TMPRSS2	0 +	41802859	41899656	0	2	1658,2123	0,94674
	chr21	41897533	42062049	ERBS4_RIPK4	0 +	41897533	42062049	0	2	2123,854	0,163662
	chr21	41897533	42608421	ERBS4_TFF3	0 +	41897533	42608421	0	2	2123,573	0,710315
	chr21	41897533	42659803	ERBS4_TFF1	0 +	41897533	42659803	0	2	2123,119	0,762151
	chr21	41897533	42689569	ERBS4_TMPRSS3	0 +	41897533	42689569	0	2	2123,986	0,79105
	chr21	41897533	42697413	ERBS4_UBASH3A	0 +	41897533	42697413	0	2	2123,401	0,79948
	chr21	41971392	42689569	ERBS5_TMPRSS3	0 +	41971392	42689569	0	2	740,986	0,717191
	chr21	41802859	41975520	ERBS6_TMPRSS2	0 +	41802859	41975520	0	2	1658,1711	0,17095
	chr21	41973809	42062049	ERBS6_RIPK4	0 +	41973809	42062049	0	2	1711,854	0,87386
	chr21	41973809	42608421	ERBS6_TFF3	0 +	41973809	42608421	0	2	1711,573	0,634039
	chr21	41973809	42643773	ERBS6_TFF2	0 +	41973809	42643773	0	2	1711,1891	0,668073
	chr21	41973809	42659803	ERBS6_TFF1	0 +	41973809	42659803	0	2	1711,119	0,685875
	chr21	41973809	42689569	ERBS6_TMPRSS3	0 +	41973809	42689569	0	2	1711,986	0,714774
	chr21	41992175	42659803	comERBS1_TFF1	0 +	41992175	42659803	0	2	2185,119	0,667509
	chr21	42250795	42608421	ERBS8_TFF3	0 +	42250795	42608421	0	2	1407,573	0,357053
	chr21	42250795	42643773	ERBS8_TFF2	0 +	42250795	42643773	0	2	1407,1891	0,391087
	chr21	42250795	42659803	ERBS8_TFF1	0 +	42250795	42659803	0	2	1407,119	0,408889
	chr21	42250795	42689569	ERBS8_TMPRSS3	0 +	42250795	42689569	0	2	1407,986	0,437788
	chr21	42348070	42643773	ERBS9_TFF2	0 +	42348070	42643773	0	2	980,1891	0,293812
	chr21	42352785	42659803	ERBS10_TFF1	0 +	42352785	42659803	0	2	571,119	0,306899
	chr21	42353360	42689569	ERBS10_TMPRSS3	0 +	42353360	42689569	0	2	470,986	0,335223
	chr21	41802859	42379590	ERBS11_TMPRSS2	0 +	41802859	42379590	0	2	1658,92	0,576639
	chr21	42062049	42379590	ERBS11_RIPK4	0 +	42062049	42379590	0	2	854,92	0,317449
	chr21	42379498	42608421	ERBS11_TFF3	0 +	42379498	42608421	0	2	92,573	0,22835
	chr21	42379498	42689569	ERBS11_TMPRSS3	0 +	42379498	42689569	0	2	92,986	0,309085
	chr21	42413539	42659803	ERBS12_TFF1	0 +	42413539	42659803	0	2	415,119	0,246145
	chr21	42631094	42643773	ERBS13_TFF2	0 +	42631094	42643773	0	2	271,1891	0,10788
	chr21	42631094	42659803	ERBS13_TFF1	0 +	42631094	42659803	0	2	271,119	0,2859
	chr21	42643773	42657235	ERBS14_TFF2	0 +	42643773	42657235	0	2	1891,878	0,12584
	chr21	42656357	42659803	ERBS14_TFF1	0 +	42656357	42659803	0	2	878,119	0,3327
	chr21	41802859	42659922	comERBS2_TMPRSS2	0 +	41802859	42659922	0	2	1658,119	0,856944
	chr21	42608421	42659922	comERBS2_TFF3	0 +	42608421	42659922	0	2	573,119	0,51382
	chr21	42643773	42659922	comERBS2_TFF2	0 +	42643773	42659922	0	2	1891,119	0,1603
	chr21	42659803	42689569	comERBS2_TMPRSS3	0 +	42659803	42689569	0	2	119,986	0,2878
	chr21	42659803	42697413	comERBS2_UBASH3A	0 +	42659803	42697413	0	2	119,401	0,3721
	chr21	42608421	42670401	ERBS16_TFF3	0 +	42608421	42670401	0	2	573,592	0,61388
	chr21	42643773	42670401	ERBS16_TFF2	0 +	42643773	42670401	0	2	1891,592	0,26036
	chr21	42659803	42670401	ERBS16_TFF1	0 +	42659803	42670401	0	2	119,592	0,10006
	chr21	42669546	42689569	ERBS16_TMPRSS3	0 +	42669546	42689569	0	2	259,986	0,19037
	chr21	42669809	42697413	ERBS16_UBASH3A	0 +	42669809	42697413	0	2	592,401	0,27204
	chr21	42659803	42810088	ERBS17_TFF1	0 +	42659803	42810088	0	2	119,909	0,149376
	chr21	42608421	43300702	ERBS18_TFF3	0 +	42608421	43300702	0	2	573,1124	0,691157
	chr21	42659803	43300702	ERBS18_TFF1	0 +	42659803	43300702	0	2	119,1124	0,639775
	chr21	42689569	43300702	ERBS18_TMPRSS3	0 +	42689569	43300702	0	2	986,1124	0,610009
	chr21	42608421	43533159	ERBS19_TFF3	0 +	42608421	43533159	0	2	573,653	0,924085
	chr21	42689569	43533159	ERBS19_TMPRSS3	0 +	42689569	43533159	0	2	986,653	0,842937
	chr21	42608421	43588574	ERBS20_TFF3	0 +	42608421	43588574	0	2	573,1372	0,978781
	chr21	42643773	43588574	ERBS20_TFF2	0 +	42643773	43588574	0	2	1891,1372	0,943429
	chr21	42659803	43588574	ERBS20_TFF1	0 +	42659803	43588574	0	2	119,1372	0,927399
	chr21	42608421	43642559	comERBS3_TFF3	0 +	42608421	43642559	0	2	573,2378	0,103176
	chr21	42659803	43643734	comERBS3_TFF1	0 +	42659803	43643734	0	2	119,1171	0,98276
MDA:ER	chr21	41802859	41935218	ERBS1_TMPRSS2	0 +	41802859	41935218	0	2	1658,708	0,131651
	chr21	41934510	42062049	ERBS1_RIPK4	0 +	41934510	42062049	0	2	708,854	0,126685
	chr21	41934510	42608421	ERBS1_TFF3	0 +	41934510	42608421	0	2	708,573	0,673338
	chr21	41934510	42643773	ERBS1_TFF2	0 +	41934510	42643773	0	2	708,1891	0,707372
	chr21	41934510	42659803	ERBS1_TFF1	0 +	41934510	42659803	0	2	708,119	0,725174
	chr21	41934510	42689569	ERBS1_TMPRSS3	0 +	41934510	42689569	0	2	708,986	0,754073
	chr21	41802859	41983188	ERBS2_TMPRSS2	0 +	41802859	41983188	0	2	1658,828	0,179501
	chr21	41982360	42062049	ERBS2_RIPK4	0 +	41982360	42062049	0	2	828,854	0,78835
	chr21	41982360	42643773	ERBS2_TFF2	0 +	41982360	42643773	0	2	828,1891	0,659522
	chr21	41982360	42659803	ERBS2_TFF1	0 +	41982360	42659803	0	2	828,119	0,677324
	chr21	41982360	42689569	ERBS2_TMPRSS3	0 +	41982360	42689569	0	2	828,986	0,706223
	chr21	41982360	42697413	ERBS2_UBASH3A	0 +	41982360	42697413	0	2	828,4	0,714653
	chr21	41992175	42062049	comERBS1_RIPK4	0 +	41992175	42062049	0	2	2185,854	0,6902
	chr21	41992175	42608421	comERBS1_TFF3	0 +	41992175	42608421	0	2	2185,573	0,615673
	chr21	41992175	42643773	comERBS1_TFF2	0 +	41992175	42643773	0	2	2185,1891	0,649707
	chr21	41992175	42659803	comERBS1_TFF1	0 +	41992175	42659803	0	2	2185,119	0,667509
	chr21	41992175	42689569	comERBS1_TMPRSS3	0 +	41992175	42689569	0	2	2185,986	0,696408
	chr21	41802859	42014469	ERBS4_TMPRSS2	0 +	41802859	42014469	0	2	1658,24	0,21137
	chr21	42014229	42062049	ERBS4_RIPK4	0 +	42014229	42062049	0	2	240,854	0,46966
	chr21	42014229	42608421	ERBS4_TFF3	0 +	42014229	42608421	0	2	240,573	0,593619
	chr21	42014229	42659803	ERBS4_TFF1	0 +	42014229	42659803	0	2	240,119	0,645455
	chr21	42014229	42689569	ERBS4_TMPRSS3	0 +	42014229	42689569	0	2	240,986	0,674354
	chr21	42062049	42197549	ERBS5_RIPK4	0 +	42062049	42197549	0	2	854,525	0,134975
	chr21	42197024	42643773	ERBS5_TFF2	0 +	42197024	42643773	0	2	525,1891	0,444858
	chr21	42197024	42659803	ERBS5_TFF1	0 +	42197024	42659803	0	2	525,119	0,46266
	chr21	42197024	42689569	ERBS5_TMPRSS3	0 +	42197024	42689569	0	2	525,986	0,491559
	chr21	42197024	42697413	ERBS5_UBASH3A	0 +	42197024	42697413	0	2	525,4	0,499989
	chr21	42062049	42334950	ERBS6_RIPK4	0 +	42062049	42334950	0	2	854,626	0,272275
	chr21	42334324	42659803	ERBS6_TFF1	0 +	42334324	42659803	0	2	626,119	0,32536
	chr21	42334324	42689569	ERBS6_TMPRSS3	0 +	42334324	42689569	0	2	626,986	0,354259
	chr21	42334324	42697413	ERBS6_UBASH3A	0 +	42334324	42697413	0	2	626,4	0,362689
	chr21	42062049	42457475	ERBS7_RIPK4	0 +	42062049	42457475	0	2	854,292	0,392506
	chr21	42405272	42643773	ERBS7_TFF2	0 +	42405272	42643773	0	2	884,1891	0,23661
	chr21	42405272	42659803	ERBS7_TFF1	0 +	42405272	42659803	0	2	884,119	0,254412
	chr21	42062049	42602888	ERBS8_RIPK4	0 +	42062049	42602888	0	2	854,219	0,54062
	chr21	42454555	42689569	ERBS8_TMPRSS3	0 +	42454555	42689569	0	2	2920,986	0,234028
	chr21	41802859	42659922	comERBS2_TMPRSS2	0 +	41802859	42659922	0	2	1658,119	0,856944

Table S1. (continued)

MDA::ER	chr21	42062049	42659922 comERBS2_RIPK4	0 +	42062049	42659922	0	2	854,119	0,597754
	chr21	42608421	42659922 comERBS2_TFF3	0 +	42608421	42659922	0	2	573,119	0,51382
	chr21	42643773	42659922 comERBS2_TFF2	0 +	42643773	42659922	0	2	1891,119	0,1603
	chr21	42659803	42689569 comERBS2_TMPRS3	0 +	42659803	42689569	0	2	119,986	0,2878
	chr21	42659803	42697413 comERBS2_UBASH3A	0 +	42659803	42697413	0	2	119,4	0,3721
	chr21	42659803	42686373 ERBS10_TFF1	0 +	42659803	42686373	0	2	119,506	0,26064
	chr21	42685867	42689569 ERBS10_TMPRS3	0 +	42685867	42689569	0	2	506,986	0,2716
	chr21	42685867	42697413 ERBS10_UBASH3A	0 +	42685867	42697413	0	2	506,4	0,11146
	chr21	42062049	42832080 ERBS11_RIPK4	0 +	42062049	42832080	0	2	854,513	0,769518
	chr21	42608421	42832080 ERBS11_TFF3	0 +	42608421	42832080	0	2	573,513	0,223146
	chr21	42643773	42939906 ERBS12_TFF2	0 +	42643773	42939906	0	2	1891,255	0,295878
	chr21	42689569	42939906 ERBS12_TMPRS3	0 +	42689569	42939906	0	2	986,255	0,250082
	chr21	42697413	42939906 ERBS12_UBASH3A	0 +	42697413	42939906	0	2	400,255	0,242238
	chr21	42062049	42983639 ERBS13_RIPK4	0 +	42062049	42983639	0	2	854,49	0,921541
	chr21	42608421	42983639 ERBS13_TFF3	0 +	42608421	42983639	0	2	573,49	0,375169
	chr21	42643773	43203101 ERBS14_TFF2	0 +	42643773	43203101	0	2	1891,166	0,557668
	chr21	42659803	43203101 ERBS14_TFF1	0 +	42659803	43203101	0	2	119,166	0,541638
	chr21	42608421	43503648 ERBS15_TFF3	0 +	42608421	43503648	0	2	573,551	0,894676
	chr21	42062049	43633495 ERBS16_RIPK4	0 +	42062049	43633495	0	2	854,172	0,1571274
	chr21	42608421	43633495 ERBS16_TFF3	0 +	42608421	43633495	0	2	573,172	0,1024902
	chr21	42643773	43633495 ERBS16_TFF2	0 +	42643773	43633495	0	2	1891,172	0,98955
	chr21	42659803	43633495 ERBS16_TFF1	0 +	42659803	43633495	0	2	119,172	0,97352
	chr21	42643773	43643734 comERBS3_TFF2	0 +	42643773	43643734	0	2	1891,1171	0,99879
	chr21	42659803	43643734 comERBS3_TFF1	0 +	42659803	43643734	0	2	119,1171	0,98276
MDA-MB231	chr21	41802859	41983188 ERBS2_TMPRS3	0 +	41802859	41983188	0	2	1658,828	0,179501
	chr21	41982360	42062049 ERBS2_RIPK4	0 +	41982360	42062049	0	2	828,854	0,78835
	chr21	42062049	42334950 ERBS6_RIPK4	0 +	42062049	42334950	0	2	854,626	0,272275
	chr21	42062049	42832080 ERBS11_RIPK4	0 +	42062049	42832080	0	2	854,513	0,769518
	chr21	41982360	42643773 ERBS2_TFF2	0 +	41982360	42643773	0	2	828,1891	0,659522
	chr21	42197024	42643773 ERBS5_TFF2	0 +	42197024	42643773	0	2	525,1891	0,444858
	chr21	42643773	42659922 comERBS2_TFF2	0 +	42643773	42659922	0	2	1891,119	0,1603
	chr21	42643773	43203101 ERBS14_TFF2	0 +	42643773	43203101	0	2	1891,166	0,557668
	chr21	42643773	43643734 comERBS3_TFF2	0 +	42643773	43643734	0	2	1891,1171	0,99879
	chr21	41992175	42659803 comERBS1_TFF1	0 +	41992175	42659803	0	2	2185,119	0,667509
	chr21	42014229	42659803 ERBS4_TFF1	0 +	42014229	42659803	0	2	240,119	0,645455
	chr21	42334324	42659803 ERBS6_TFF1	0 +	42334324	42659803	0	2	626,119	0,32536
	chr21	42454555	42689569 ERBS8_TMPRS3	0 +	42454555	42689569	0	2	2920,986	0,234028
	chr21	42659803	42689569 comERBS2_TMPRS3	0 +	42659803	42689569	0	2	119,986	0,2878
	chr21	41982360	42697413 ERBS2_UBASH3A	0 +	41982360	42697413	0	2	828,4	0,714653
	chr21	42334324	42697413 ERBS6_UBASH3A	0 +	42334324	42697413	0	2	626,4	0,362689
	chr21	42659803	42697413 comERBS2_UBASH3A	0 +	42659803	42697413	0	2	119,4	0,3721
	chr21	42697413	42983639 ERBS13_UBASH3A	0 +	42697413	42983639	0	2	400,49	0,286177

Table S2. Correction coefficients used to generate MCF-7 functionalized network of ERBS-genes promoters interactions.

Thanks to the data obtained using TFOs to interfere with ER binding and functions at specific places we were able to generate functionalized interactomes from the identified network of ERBS-genes promoters interactions. For that, we integrated the effects of TFOs on ER binding, interactions between ERBS and given promoters and their consequences on target genes RNA levels into a global coefficient as shown in this table by dividing unlogged values for each parameters. A value of 0 was given for undetermined data, and we arbitrarily set values to 1 (division of the final coefficient by a factor of 2) when the disruption of ER binding at specific ERBS had no significant consequences on either three-dimensional interactions or RNA levels.

Origin	Target	Correction for TFO mediated drop in ER	Correction for TFO mediated drop in interaction	Correction for TFO mediated direct effect on RNA levels	Final correction coefficient
		A	B	C	$-2^A/2^B/2^C$
comERBS1	TFF1	-0.603040604915369	-2.31344638198086	-0.513337106747387	4.67103703503524
comERBS3	TFF1	0	0	0	1
comERBS3	TFF3	0	0	0	1
ERBS1	UBASH3A	-1.55329362553128	-5.39896913065119	-0.999134642556747	28.7364540924603
ERBS10	TMPRSS3	-0.659386488075739	-3.01391678033519	1	2.55713848256888
ERBS10	TFF1	-0.659386488075739	-1.44929630375211	-0.27299444968944	2.08913289004776
ERBS11	RIPK4	0	0	0	1
ERBS11	TFF3	0	0	0	1
ERBS11	TMPRSS2	0	0	0	1
ERBS11	TMPRSS3	0	0	0	1
ERBS12	TFF1	0	0	0	1
ERBS13	TFF1	0	0	0	1
ERBS13	TFF2	0	0	0	1
ERBS14	TFF2	-0.789097648704165	-1.69059095075988	1	0.933999253553983
ERBS14	TFF1	-0.789097648704165	-2.05050241879443	1	1.19864527237376
ERBS16	TFF1	-1.43495208386755	-1.1814766179337	-0.571535973742021	1.24665347884363
ERBS16	TFF2	-1.43495208386755	-1.79636030598601	-0.434987819101775	1.73675715540747
ERBS16	TFF3	-1.43495208386755	-1.48883561141615	1	0.519027731147108
ERBS16	TMPRSS3	-1.43495208386755	-1.30080375485176	0.24061962931176	0.771229446525847
ERBS16	UBASH3A	-1.43495208386755	1	1	0.0924647903038572
ERBS17	TFF1	0	0	0	1
ERBS18	TFF1	0	0	0	1
ERBS18	TFF3	0	0	0	1
ERBS18	TMPRSS3	0	0	0	1
ERBS19	TFF3	0	0	0	1
ERBS19	TMPRSS3	0	0	0	1
ERBS2	TFF3	0	0	0	1
ERBS2	TMPRSS3	0	0	0	1
ERBS20	TFF1	0	0	0	1
ERBS20	TFF2	0	0	0	1
ERBS20	TFF3	0	0	0	1
ERBS3	RIPK4	0	0	0	1
ERBS3	TFF2	0	0	0	1
ERBS4	TFF1	-1.24490279856471	-0.879648059634778	-0.242635327614509	0.918518439156855
ERBS4	TFF3	-1.24490279856471	-1.1427355549934	-0.737446572680709	1.55323845353845
ERBS4	TMPRSS2	-1.24490279856471	-2.25704524330251	-2.26257281727094	9.67804241515562
ERBS4	TMPRSS3	-1.24490279856471	-0.886832943267265	1	0.390103850800125
ERBS4	UBASH3A	-1.24490279856471	1	1	0.105484081256354
ERBS5	TMPRSS3	0	0	0	1
ERBS6	RIPK4	-2.00642468829735	-0.772751816219267	1	0.21261674492913
ERBS6	TFF1	-2.00642468829735	-0.865384755574892	-0.384018789185219	0.591716838857362
ERBS6	TFF2	-2.00642468829735	-1.58207999218803	-0.451859343104953	1.01925472655034
ERBS6	TFF3	-2.00642468829735	-2.07400058144378	-0.78687049999252	1.80806481140562
ERBS6	TMPRSS2	-2.00642468829735	-0.935668953754016	-0.691056711874786	0.768597920038495
ERBS6	TMPRSS3	-2.00642468829735	1	1	0.06222229040351
ERBS8	TFF1	0	0	0	1
ERBS8	TFF2	0	0	0	1
ERBS8	TFF3	0	0	0	1
ERBS8	TMPRSS3	0	0	0	1
ERBS9	TFF2	0	0	0	1
TFF1	TFF2	-1.49188850724005	-1.06793882865658	-0.582295980666751	1.11600717465538
TFF1	TFF3	-1.49188850724005	1	-0.624821520410388	0.274130168512921
TFF1	TMPRSS2	-1.49188850724005	-1.976531719	1	0.699619890670991
TFF1	TMPRSS3	-1.49188850724005	-1.21591485726524	1	0.412946374635607
TFF1	UBASH3A	-1.49188850724005	-2.65926683941227	1	1.12301588301092
TFF1	TFF1	-1.49188850724005	0	-0.461163892258534	0.489464246367371

Table S3. Correction coefficients used to generate MDA::ER functionalized network of ERBS-genes promoters interactions.

Thanks to the data obtained using TFOs to interfere with ER binding and functions at specific places we were able to generate functionalized interactomes from the identified network of ERBS-genes promoters interactions. For that, we integrated the effects of TFOs on ER binding, interactions between ERBS and given promoters and their consequences on target genes RNA levels into a global coefficient as shown in this table by dividing unlogged values for each parameters. A value of 0 was given for undetermined data, and we arbitrarily set values to 1 (division of the final coefficient by a factor of 2) when the disruption of ER binding at specific ERBS had no significant consequences on either three-dimensional interactions or RNA levels.

Origin	Target	Correction for TFO mediated drop in ER	Correction for TFO mediated drop in interaction	Correction for TFO mediated direct effect on RNA levels	Final correction coefficient
		A	B	C	$=2^A/2^B/2^C$
comERBS1	RIPK4	-0.861179459257291	1	-0.676918210496267	0.44004981365524
comERBS1	TFF1	-0.861179459257291	-1.81326116051221	-0.647693447093664	3.03096070495228
comERBS1	TFF2	-0.861179459257291	-0.922179195279462	1	0.521594201325967
comERBS1	TFF3	-0.861179459257291	-1.61549438321474	-2.28982725172034	8.24855939486474
comERBS1	TMPRSS3	-0.861179459257291	0.444449645205835	1	0.20227233162859
comERBS3	TFF1	0	0	0	1
comERBS3	TFF2	0	0	0	1
ERBS1	RIPK4	-1.44720517162484	-3.06381592091033	-0.524292505357361	4.41038088763461
ERBS1	TFF1	-1.44720517162484	-2.59193522526698	-4.20822759572263	40.8696405392146
ERBS1	TFF2	-1.44720517162484	-2.40711346155891	-0.814006629930034	3.41980702914851
ERBS1	TFF3	-1.44720517162484	-1.78024910240811	1	0.629834168446143
ERBS1	TMPRSS2	-1.44720517162484	-2.08505343210263	1	0.778003343554803
ERBS1	TMPRSS3	-1.44720517162484	-1.17018191171104	-1.84114597048338	2.95697636186825
ERBS10	TFF1	-1.75939889525682	-1.44149343374557	-0.312043505955518	0.995945045653898
ERBS10	TMPRSS3	-1.75939889525682	-1.83256780618123	-0.22398025935253	1.22871401916829
ERBS10	UBASH3A	-1.75939889525682	1	1	0.073842801833693
ERBS11	RIPK4	0	0	0	1
ERBS11	TFF3	0	0	0	1
ERBS12	TFF2	-1.19502468906442	-3.03732419486704	-1.13637866025779	7.88263605441031
ERBS12	TMPRSS3	-1.19502468906442	-5.09040276871721	1	7.44038914986413
ERBS12	UBASH3A	-1.19502468906442	-1.10952840583346	1	0.471230138168896
ERBS13	RIPK4	0	0	0	1
ERBS13	TFF3	0	0	0	1
ERBS14	TFF1	0	0	0	1
ERBS14	TFF2	0	0	0	1
ERBS15	TFF3	0	0	0	1
ERBS16	RIPK4	0	0	0	1
ERBS16	TFF1	0	0	0	1
ERBS16	TFF2	0	0	0	1
ERBS16	TFF3	0	0	0	1
ERBS2	RIPK4	0	0	0	1
ERBS2	TFF1	0	0	0	1
ERBS2	TFF2	0	0	0	1
ERBS2	TMPRSS2	0	0	0	1
ERBS2	TMPRSS3	0	0	0	1
ERBS2	UBASH3A	0	0	0	1
ERBS4	RIPK4	0	0	0	1
ERBS4	TFF1	0	0	0	1
ERBS4	TFF3	0	0	0	1
ERBS4	TMPRSS2	0	0	0	1
ERBS4	TMPRSS3	0	0	0	1
ERBS5	RIPK4	0	0	0	1
ERBS5	TFF1	0	0	0	1
ERBS5	TFF2	0	0	0	1
ERBS5	TMPRSS3	0	0	0	1
ERBS5	UBASH3A	0	0	0	1
ERBS6	RIPK4	0	0	0	1
ERBS6	TFF1	0	0	0	1
ERBS6	TMPRSS3	0	0	0	1
ERBS6	UBASH3A	0	0	0	1
ERBS7	RIPK4	0	0	0	1
ERBS7	TFF1	0	0	0	1
ERBS7	TFF2	0	0	0	1
ERBS8	RIPK4	0	0	0	1
ERBS8	TMPRSS3	0	0	0	1
TFF1	RIPK4	-0.87151982189975	-1.71755243387244	-1.01887801020569	3.64245367014246
TFF1	TFF2	-0.87151982189975	-1.76621328870842	-1.3122226226037	4.61687313389505
TFF1	TFF3	-0.87151982189975	1	0	0.273285378331784
TFF1	TMPRSS2	-0.87151982189975	-1.26466117621385	-3.40506933018761	13.9115443906158
TFF1	TMPRSS3	-0.87151982189975	-3.27437574461058	-0.804652401677781	9.2375375541922
TFF1	UBASH3A	-0.87151982189975	-2.62608621909348	-1.03386489456205	6.90878153091822
TFF1	TFF1	-0.87151982189975	0	-0.910773057929605	1.02758179481776

Table S4. Antibodies used.

Multiple references are cited for the same protein and application when mixes of antibodies were used.

ChIP: chromatin immunoprecipitation; MeDIP: methylated DNA immunoprecipitation; WB: Western blot.

Target	Reference	Provider	Application
5mC	MAB-006-100	Diagenode	MeDIP
CTCF	07-729	Millipore	ChIP, WB
ER	HC20	Santa Cruz	ChIP, WB
ER	ab 10 (TE111-SD11)	NeoMarker	ChIP
FOXA1	ab 5089	Abcam	ChIP
FOXA1	ab 23738	Abcam	WB
GATA2	ab 22849-100	Abcam	ChIP
GATA2	sc 9008	Santa Cruz	ChIP, WB
GATA3	sc 22206	Santa Cruz	ChIP, WB
H3	ab 1791	Abcam	ChIP
H3K4me1	07-436	Millipore	ChIP
H3K4me2	07-030	Millipore	ChIP
H3K14ac	06-911	Millipore	ChIP
H3R17me2	07-214	Millipore	ChIP
H3K27ac	Ab4729	Abcam	ChIP
Pol II	05-623 (CTD4H8)	Millipore	ChIP
Pol II	sc 13583	Santa Cruz	ChIP
RAD21	ab 992	Abcam	ChIP
Smc3	/	Dr. JM Peters	ChIP, WB

Table S5. Recombination cloning primers.

	Name	5' oligonucleotide	3' oligonucleotide
MDA::ER	ERBS1	ATTAAAGGAATTCCTGCAGGCAGCAGAGATTGAAAAGGT	GCTCTTCTCCACTAGTGATCAAGGGGTCATTGTTC
	ERBS2	ATTAAAGGAATTCCTGCAGAATAACAGGGCAGCCTCCTA	GCTCTTCTCCACTAGTCCCAACCATCTAGGAAATG
	ERBS14	ATTAAAGGAATTCCTGCAGCCTGAGCTCAAGCAATCCTC	GCTCTTCTCCACTAGTCTGGGAAACAGAGACTCCA
	ERBS15	ATTAAAGGAATTCCTGCAGGTAAAGCTGGCTGCATACGA	GCTCTTCTCCACTAGTCGGGTGAGGTTGTATGCTAA
MCF-7	ERBS4	ATTAAAGGAATTCCTGCAGGTCCTGGCTTCTGAGTGAGG	GCTCTTCTCCACTAGTCATCTGTAAAGGCCCTGCTC
	ERBS5	ATTAAAGGAATTCCTGCAGAACAGGAAACCGTTGAGGTG	GCTCTTCTCCACTAGTCCCACTGGCAGATAGAGAA
	ERBS9	ATTAAAGGAATTCCTGCAGGTCACATGGCTGGTTGACAC	GCTCTTCTCCACTAGTCAGTCACTCAGGGGTTTGGT
	ERBS14	ATTAAAGGAATTCCTGCAGGGGAGGGAGAAAGAGATGCT	GCTCTTCTCCACTAGTGACACAAATGGTGCTCAGT
	ERBS16	ATTAAAGGAATTCCTGCAGGGGACCTTAGGGTCAAAAGC	GCTCTTCTCCACTAGTGCTTCGAGACAGTGGGAGTC
	ERBS17	ATTAAAGGAATTCCTGCAGTGCTCTAATGGGCTGTTTCC	GCTCTTCTCCACTAGTCCACGTGCTCAAGAACTCA
	ERBS20	ATTAAAGGAATTCCTGCAGAGTTGGCATGAAGGATCAGG	GCTCTTCTCCACTAGTCGTGTGTGGGATGTTGACTC
Both cells	comERBS1	ATTAAAGGAATTCCTGCAGTGAAGGGCTTTCATTGTG	GCTCTTCTCCACTAGTGAATCTGAAGCTGCCCTTTG
	comERBS2	ATTAAAGGAATTCCTGCAGGGAGATGTTGGCATGAACAG	GCTCTTCTCCACTAGTCCATAGACAGGGGAGTGA

Table S6. siRNA sequences.

Target	siRNA sequence (sense strand)
Luciferase	AACACUUACGCUGAGUACUUCGA
CTCF	GGAGCCUGCCGUAGAAAUU
RAD21	GGUGAAAAUGGCAUUACGG
GATA2	GGCUCGUUCCUGUUCAGAATT
GATA3a	AAGCCUAAACGCGAUGGAUUA
GATA3b	CCGAACUGUUGUAUAAAUUUA
GATA3c	GAGGUGGAUGUGCUUUUUAAC
FOXA1	GGACUUCAAGGCAUACGAAUU

Table S7. RT-qPCR primers.

Target RNA	Forward primer	Reverse primer
<i>Rplp0</i>	GCAAGAACACCATGATGCGC	CTGATCTCAGTGAGGTCCTC
<i>CTCF</i>	TGTGGCAGGGCATTGAGAACAG	ACTTGTGAGGACGAGTACCTGTG
<i>Sccl</i>	AGGTTCTAGAGAGGACCAGCAAC	TCCTAGAGAGGACCAGCAACAG
<i>GATA1</i>	CACAAGATGAATGGGCAGAACAGG	CCTGCCCGTTTACTGACAATCAGG
<i>GATA2</i>	GCAAGGCTCGTTCCTGTTGAGAAG	TGGTCGGTCTGCCCATTTCATC
<i>GATA3</i>	TTAAGCCCAAGCGAAGGCTGTCTG	GTGGTTGTGGTGGTCTGACAGTTC
<i>FOXA1</i>	GAAGATGGAAGGGCATGAAA	GCCTGAGTTCATGTTGCTGA
<i>ER</i>	TGGAGATCTTCGACATGCTG	TCCAGAGACTTCAGGGTGCT
<i>TMPRSS2</i>	TGTGGTCCCTTCCAATGCTGTG	TGCTCATGGTTATGGCACTTGGC
<i>RIPK4</i>	CCATGTCCACTGGAAGACCT	CCATCTCCATCTTCTTGCT
<i>TFF3</i>	CTCCAGCTCTGCTGAGGAGT	CACTCCTGGGGGTGACA
<i>TFF2</i>	ATGGATGCTGTTTCGACTCC	TCTGAGACCTCCATGACGC
<i>TFF1</i>	GGAGAACAAGGTGATCTGCG	AATTCTGTCTTTCACGGGGG
<i>TMPRSS3</i>	ACTGGATCCACGAGCAGATG	GCTGTCTTCATCACCTCAGG
<i>UBASH3A</i>	TTCCTGCCGGAATACTACAC	TGGAGTTCAGGTCTGTGGCT
<i>PKNOX1</i>	TCAGTGTGATGAAGATTGGCA	AGCCATCATGGTTCATCAGA
<i>EML2</i>	TGGGTACAGTGACTGGCAGATG	ATTGCCGTCTGTGTGGATAGC
<i>CDH18</i>	AGGTGAAACCGAAGTCCATCATCG	TGTCAGAATTGGAGTGAGCTTTC
<i>VAV2</i>	GATCCACATCACAGAGGCCAAG	TGGTAGTACTCCACCAACTCCAG
<i>FAM210A</i>	AAGCGGGCATGTGAAGGAAG	CAGTCCGTGGTACATTCCATTGC
<i>COL14A1</i>	TGTTCTGCCTATGGTGTGAGAGC	TGGCTTCAGCTCATCTTGAC
<i>GPRASP1</i>	ATCCAACTGAAGAGCAGCCAGAG	CCTGTCAGAATACGCCAGACCTC
<i>TNPO2</i>	TCCATGTCAAGCCCTGTATCGC	TGCAGACGGAGATGAACTCTGG
<i>WDR74</i>	CCAGGATCACAGAAGCTTGTACCC	TGCTGGATCATAAACACGGACCTG
<i>TNPO1</i>	AAGACTCTTGACAAGCCATGC	TGTTGCCTCCAAGTCCTTCAGC
<i>TNNI3K</i>	GAGTGTCTCTGCAACATTGAGCTG	AAGAAGGTGAGAGAGACCCACTGC
<i>HSPA9</i>	ACCAATTACCTGCTGATGAGTGC	ATGCCTGCTGAAGAGAGGATGC
<i>SNAP29</i>	ACCCAAAGAACCCACACCTTCG	AGTCCCATGGACAGCTCATCTAGG
<i>DDX26B</i>	GCGCTGTGGAGTTATTCTGAAGC	ACCAGCATGTACCTGTCTCCAC
<i>MLLT10</i>	TCTCAACAGCTCACACCAGTACAC	TGGATGGTGAGAAATGGGTTGGTG
<i>PDCL</i>	TGGATGAAGAGGAGGAGCAACAG	TCCTTCAGAGTCATCTTCCACTG
<i>CHD5</i>	ACCACACGCTGACCAACTACAAG	TCTTGGCAATGAGTGGCCTGAG
<i>TLE3</i>	CCTGCAAGCTCAGTATCACAGC	TGCATCTCCGTCTTCTCGTTTGC
<i>NPR3</i>	GCTCTCCCTGCAATCATGTGG	CAAGCCAGCTCCTAGTAAAGCC
<i>NR2F2</i>	TGGTCGCCTTTATGGACCACATAC	TCAGAGAGACCACAGGCATCTGAG
<i>BDNF</i>	TGGGAAGAAGGAGAACTGAAGCC	CTGGTGGAAGTGAAGATTAGATGGC
<i>DLGAP1</i>	AACAACCTGCCCAGAACATTC	ATTTCTGGGCCATGAGAAGTTGGG
<i>ZNF841</i>	GCCTACAAGTGGTCCTTACACCAG	TGAGGTGTGCGATTGAGTTGCC
<i>AFF4</i>	AGCCATCCAGCAACAGAGAGTC	AGCGGGAGAAGGCAACAAATCC
<i>AFG3L2</i>	AAGCTGACGTGGAGAAGGTTGC	TGCCAGTGCCTTCCACAAATTC
<i>PNPLA4</i>	CGACTAAGAAGTGGGATGGAGTCG	TGGATACGTGCAGTCGGTTCTG
<i>SHPRH</i>	GGCGAGAGATTGTTACATCTGAGG	TTGGGTACTCACGAATGATGCAG
<i>HAUS6/FAM2</i>	GGCGAGAGATTGTTACATCTGAGG	TTGGGTACTCACGAATGATGCAG
<i>TLL1</i>	AGACAGAAGGAGCTGTGAAGCTG	GGGTGGTTATGGTGCCGTTAAG
<i>PTPN13</i>	ATGAAGTGCTCAGAGCCAACC	ACATCCTCACACATTCCAAGCAG
<i>COPB2</i>	ACTTACCCAGTCAGGTTTCAAGGG	TTGGGTGAGCAAGGGATTCTGC
<i>PPARGC1B</i>	AGGCGCTTTGAAGTGTGTTGGTG	ATGAAGCCGTACTTCTCGCCTCTC

S100PBP	CCCAGAGGACACTAACCAAGGATG	TGCTGGGATTACAGGCATGAGC
SF3A1	ATGATGAGGTGTACGCACCAGGTC	AAGATGTCAGTACGCCGCTCAG
ZNF354A	AGAAAGACGGTTCTGGCGTCTC	TGTGTTTGCGTTGACTTTGTGG
MAP4K4	AGGACAGAGAAACTCCACCAGCAG	TTCTCCACTCTCTCCCACAGAAGC
CD99L2	ACCACCACCACAACCAATAGGC	AGCCAAGTCCAATCCACTACCTG
NPY2R	CACATCTTGTTTCCGCGTCTCC	AACCAGCACAAAGAGTCTACAACCTG
MBD1	CGCTCAGACACCTATTACCAGAGC	TCGAGTCAGCTCAACTTTGCTTCG
LILRA3	TGCAGGGCGGTATTGCTGTATC	TTTGCTGTAGGCTCCTGTCACC
TAF2	GCACGGAAGGGAACAGGTAAAG	AAGACTGAGAGTGGAGCGCTTG
OTP	AACGAGTTGGAGAGGAGCTTCG	TTCTGGAACCAGACCTGCACTC
LILRA1	AAGTCCGATTCCAAGGCTGGAG	GCTATGCCCATGCGGATGAGATTC
EFEMP1	GTGAACACAACGTGTGCCAAGAC	TGATGCACACTTGGTCTGCTCTAC
DICER1	ACGGTGTTCTTGGTCAACTCTGC	AGTTCTGACAGCTGACACTTGTTG
RALGPS1	GCTGAACAACCCTGACAAAGGC	TGCATGAAATCGGGAACCAGTC
SGCD	TGCCTTCACCCAGAGCTTAGAG	AGCACATCAGAGAGGGATGGAC

Table S8. ChIP-qPCR primers

target type	target name	Forward primer	Reverse primer
controls	Rplp0p	ATCTAACTAGCACACGAACCTT	CTGTATTCGTTTCTTGTCT
	CXCL12p	CACCATTGAGAGGTCGGAAG	AATGAGACCCGCTTTGCAG
	PKNOXp	GGGTTACGGATGGTACAGC	CCTCCCGCCTTATGACTATG
MCF7 ER binding sites (ERBS)	ERBS1	GTGGGAATGGAGGACTAGCA	TCCAAGGCCAAGTTCTTGAT
	ERBS2	CGGGCGTTATATGGACTCAT	ATCACCAAATCCCCACGTTA
	ERBS3	CATGACCTTGACGCTGAGAA	GGAGGATGGGACATTTGTTG
	ERBS4	GAGCCCTGTAACTGATGGA	TGGTGTCTTCTCCCGTATC
	ERBS5	AGGGAGCAGCCTGTCTGTAA	ATCAGTGCTCAGCTCCCAGT
	ERBS6	GACACCTGCTGTTCCCTCAT	CCAGAGAGAAGCCACTGACC
	ERBS8	ACCCTGTGGGCACATGTTAT	CCCTGGTTGACTTCACTGGT
	ERBS9	AGAACGGAAGTGCAGGAAGA	AGAACCCAAACCCAGTGATG
	ERBS10	GTCGGCCTCTTCTGATTCTG	CCCAGGACAATGTAGGTGCT
	ERBS11	CTGGAAATGCCGAACACTTT	GTCTAACGGCCAAGGAGTGA
	ERBS12	AGTCTGCTGTGGCTGGAGTT	CGTCCTCTTCTTTTACAC
	ERBS13	AAGCTGCGTGTTGTCAGATG	AAGGAAGGGCCCTCTATTCA
	ERBS14	TGACACCAGGAAAACCACAA	CCTCACTGGACAGTTGCTGA
	ERBS16	CGTCACCTGCTTCAAAGTGA	GCTTCGAGACAGTGGGAGTC
	ERBS17	GAATCTGCCCCACTTATCCA	AAGACCTCTGGAAGCACGAA
	ERBS18	GGAAGCTGAGGTCAGGACAG	AACAAGCAATGCCTTTGTCA
	ERBS19	CAGGTGGGTGTGTGCAGTAG	ACCCAGATGTCCTTACCAG
	ERBS20	GCTGCTGTACTGTGCTGACC	CAATTCAGAGGCAGGGTTGT
MDA::ER ERBS	ERBS1	CTAAACCTGGCAGAGGTTGC	GAGGGACTGAGGTCATTGGA
	ERBS2	GCTGCTTCAGGTAGCTTTGG	AGTTAGGAGGCTGCCCTGTT
	ERBS4	CAACAGGGTGTTGGGGTAAC	ATCACCCACCAGCTAAGTGC
	ERBS5	CAAGACGGCAATGAGAGTGA	GTGCTGGTGGGAGTGCTTTT
	ERBS6	CCTGCCTGTCTTTCTGAAG	CTCCTGGGACCTTACCTTCC
	ERBS7	TGCAGATGCTCTGGATTGTC	GGTGCAGGAACCTCTGGATGT
	ERBS8	TGGGAGGAATTGCTTTTGAC	TCAATCAAGGAGCTCCACCT
	ERBS10	AGACACCGGAAAGCTCAAAA	GTCCTTGGGGGTGATATTT
	ERBS11	CCCCAGCATTTTACACCTTA	TTTCAGCTGCTTTGCCACTA
	ERBS12	GAGAAAGATCCCTGGGGAAG	AGTGGACAGCAAGTGCCTCT
	ERBS13	GAAACCAGCACCTTGCCAT	GGTGGGCTTCTCTATTTC
	ERBS14	AGCTCAGGAAATCGAAACCA	CCACCTCCCAGCTTCAAGT
	ERBS15	CAAAGGCCTCTGAGTTCTGG	TGTGACAACCCACAGCTCTC
	ERBS16	CAATCATGGGGGTTCAAAAC	ACCTGCCTCAAGACGGTGAC
common ERBS	comERBS1	GGCTTTCCATTTGTCTCTGC	GGAAATGCAAGACCAGCAAT
	comERBS2	GTCGCACTTCTCGAAGGTCT	CAGGTAAGGCGTGCTTCTTC
	comERBS3	CACCAGGTCACCACCCTACT	GGAAGGGGTGGCTTCTGG

Table S9. Primers for amplification of circularized 4C products.

Anchored promoter	Primer 1	Primer 2
TMPRSS2	AACATAGTCCTCTTTGGCACA	GTCAGTCTCGGGGAGGGACT
RIPK4	TTGGGAGCTTCCATCAAGAC	GCTCCTTCATGGGTTCATTC
TFF3	GACCAGGGTGTGGTGTCC	CAGCTCTGCTGAGGAGTACG
TFF2	CAGACCCTCATCCTCCAGAC	TATAAAGGCATCTGGCAATGTG
TFF1	GCTACATGGAAGGATTTGCTG	CAGTGGAGATTATTGTCTCAGA
TMPRSS3	CATGGCTGCTCTGGGAAC	CCTCGGCTAAGGAGGTAGAG
UBASH3A	GTACGGCTTCCTGCCAAA	CCGCTGCCATCTCTTCCT

Table S10. Primers for 3C/4C fragments amplification and controls.

		Forward primer	Reverse primer
Controls	Digestion and 3C/4C internal normalization	AAGTGTGGTGGGAAGGACAG	GGACCAAGGATTTGGGTTT
		3C primer(s)	Digestion primer
	Positive (GAPDH proximal ligation)	AAGACAGAATGGAAGAAATGTGC GGAGCCAAAAGGGTCATCAT	CGAGATCCCTCCAAAATCAAG
	Positive (TFF2 proximal ligation)	CTCTGTGCCTATGCCATCTG	TGCAAAATGGAAGTGAGCAG
	Negative (GAPDH to X)		
	Negative	CTCAGCCTTCGTTTGCATAG	GGCCTGGATACAGGATCTTC
	Fragment	3C primer	Digestion primer
Promoters anchors	TMPRSS2 RIPK4 TFF3 TFF2 TFF1 TMPRSS3 UBASH3A	GAGGAAGTGCAGACACCAGA CATCTGTGTGCTTCATTTGACA AGCCAGAAGCCCTCAGAAC CTCCAGGCTCTCTGCTGACT GCTACATGGAAGGATTTGCTG TCCTCCAAGGGCTAAAATCC AGTGACACCTGCAGGGTAAG	GCCGTGTGAGGCAGATAAAA CTTCAAAGTCCGCAGAGGAG AGGGGAGAAAGCAAAAAGGA AGGAGAGGTGGGTGGGAATA TTCCGGCCATCTCTCACTAT AGTCACAAAGGGAATGAAAGC TTAAGCCCAAGCTGTCCTG
MCF7 ER binding sites (ERBS)	ERBS1 ERBS2 ERBS3 ERBS4 ERBS5 ERBS6 ERBS8 ERBS9 ERBS10 ERBS11 ERBS12 ERBS13 ERBS14 ERBS16 ERBS17 ERBS18 ERBS19 ERBS20	CTGCCTGAGGGAGAGAAAAAG GTCCAGAGCCCAGGATGTT TAGAACCAAGCTGGCAGGTG ACAGGAGAAAGGCACGTGAA AGCCCTGAGGGTAGGAAC GGTCAGTGGCTTCTCTCTGG TTCAGTAACGAAGATATCTGCTT GCCAAACAAGAAGGGCCTA TCACAGCACCTCAGACAAGG CCATATGTCCCTGGGTGTCT TCAACAATCAACCCAGCATC CGCTCGGAAGGTTGAATAGA CATTGTGGCACACGTCTGTA GAGCATGGAGGGGGAGAT ACGTTTGCTTCTGACCATCC TTCCTGTTGGAGCATGGAAT AGTGTCTCACCAGGGCTCAG GACCAAGCAAGGGCTTCTG	TCTATTATTTTGCTTAATCTCATGCTA ATTTACGTGGCCACTGGTT GGAGAGAAGAAAATTGGGGAAA TTAGGTGATCTGCCGGTCTT CCCACTGGCCAGATAGAGAA AATAGAAGGCTCACCGTGGA AGAGAAGGCTGGTGGGATTT TTAATGGGATTTGCCTCA GTGCTGAGAGCTCTGCCTCT CGAGGCTGCAGACTACCTGT CCTTCAGTTCCTGGGAGTA TGCAGCCTCTGTGTTTGT CTTGCTTTTCTCCTCCATCA TCACTTTGAAGCAGGTGACG CAAAACCACGTCTGCTCAAA ATCGGGTTCAGAGCAGTTTG GCCTTGGTCTACGATGAGGA ATCTTGGGGTGAAGTGAAGCA
MDA::ER ERBS	ERBS1 ERBS2 ERBS4 ERBS5 ERBS6 ERBS7 ERBS8 ERBS10 ERBS11 ERBS12 ERBS13 ERBS14 ERBS15 ERBS16	AAAGTTCAAAACGGCCAGAC TGCTGCAATGGATATGAGAGA CCTGGTACCAAAAAGTTGCTTC CACTGCCAGGGGAAGAACT GCCCTGATGCCACTCTTC GCACCTCTAGTCATACCGTAACC TAGAAAGCCTGGCCTACCA GGGCATGTGCCTGTAATTC CCCTCGAGGCTGTTTATGTG AAGCACCCCTCGTGAAG TGGACGTAGAAAGAGCTTCAGAT GCTGGTACTACAGGCTCACG CGTCACGGGCTCCTACTTTA AGTCCCTCCAGAAAAATCG	GGGGATATTGTGTTCTGATG TGGGAACGTGGATCAGTACA CGGCAATATTTGGGGTAAC CCTGTGAGTAAGGGCACCTC CACACAGCAAAGCTTCCATC GTCACGGCTGAGATCCAACT ATATTCTGTGGCCTGCAACC GGACATTTCTTTCACTTCA AGTGGCTTTCAGCTGCTTTG GGGGGAGTGAAAGTGACAGA CAGTTTGTGATGAGAGAGGAGAG CAAAAACCAGGGCTGTAGGAT TGACTCAGATTTGCCTCGTG GAAGGTGCTCTGCAGAGGAG
common ERBS	comERBS1 comERBS2 comERBS3	ATCCTTGGTCCTTCCAGGAG GCTACATGGAAGGATTTGCTG CATCTGCTTGGCTGACTGG	GACTTGATCCAGGGTGGTA TTCCGGCCATCTCTCACTAT CTCACTCCCTGGCTGCAC

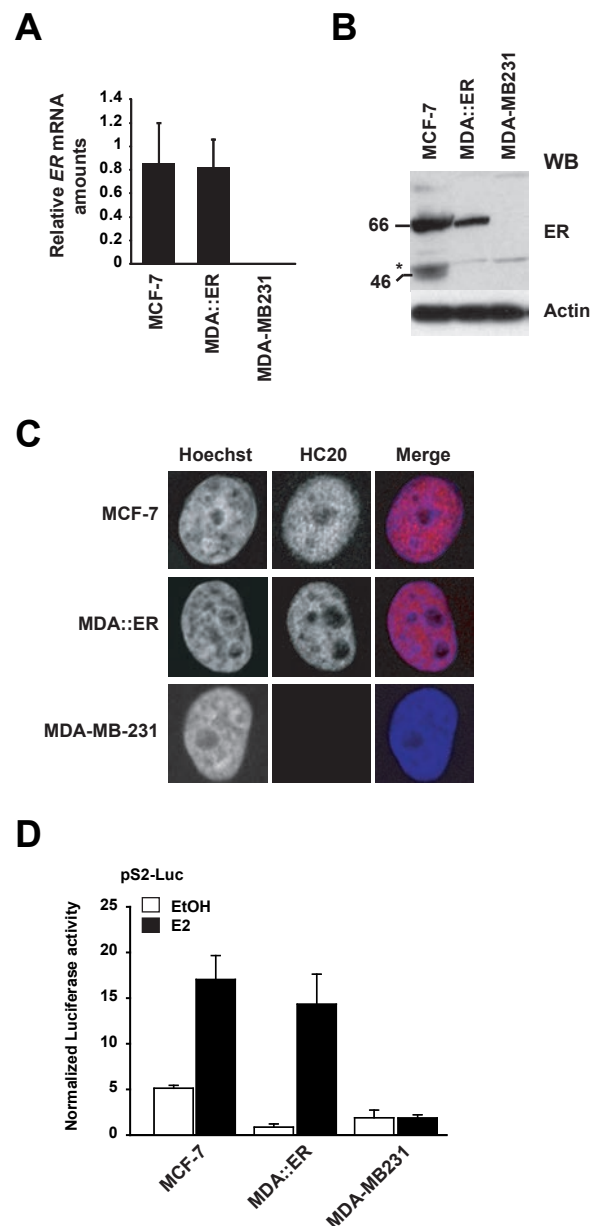


Figure S1. Related to Figure 1. Phenotypic comparison between MCF-7 and MDA::ER cells.

(A) RT-qPCR quantification of *ER* expression in MCF-7, MDA::ER and MDA-MB231 cells. Results were normalized to the expression of the *Rplp0* control gene.

(B) Western blot analysis of ER expression in all three cell lines. Actin was probed (separate gel) as an input control.

(C) Confocal immuno-fluorescence images of ER in MCF-7, MDA::ER and MDA-MB231 cells labeled with HC20 antibody, and visualized with a secondary antibody coupled to Cy3.

(D) A Luciferase reporter placed under the control of the pS2 promoter (-712/+30) was transiently transfected in the three cell lines for 36h. Following a 2h treatment with 10-8M estradiol (E2) or ethanol (EtOH) as vehicle control, Luciferase activity was measured and normalized to that of a control Renilla plasmid. Results show are mean \pm SD of 3 independent triplicates.

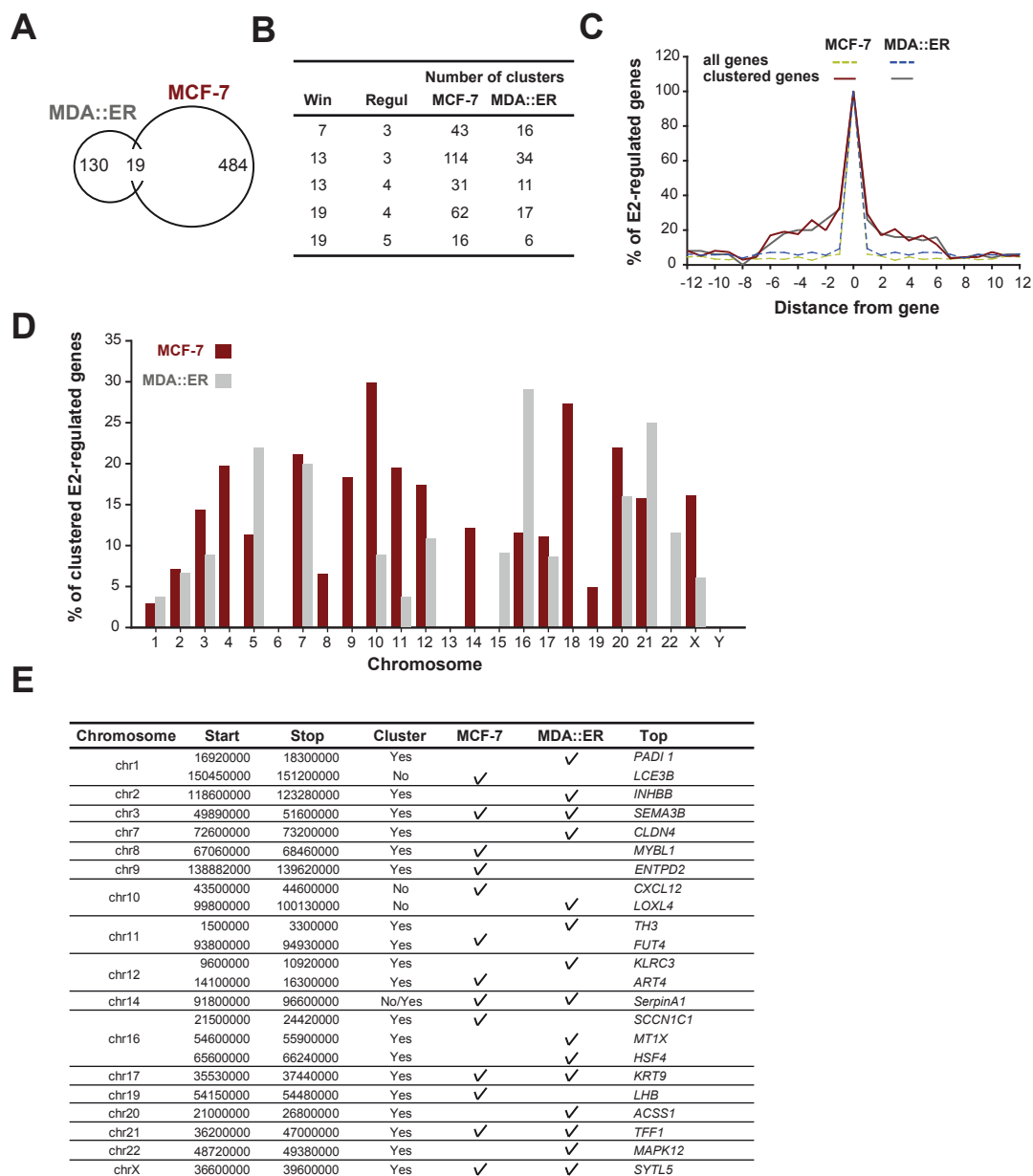


Figure S2. Related to Figure 1. E2-sensitive gene clusters within MCF-7 and MDA::ER cells.

(A) Venn diagram comparing the estrogen-sensitive genes identified in MCF-7 and MDA::ER cell lines following a 4 h treatment with 10-8M β -Estradiol (E2).

(B) To identify E2-sensitive gene clusters, we first built a comprehensive gene dataset restricted to non-redundant and unambiguously annotated (positioned) genes regulated after 4h of E2 treatment from the GSE23850. Clusters were then identified within this filtered dataset using a sliding window (win) of different length to define genomic regions containing 3 or more E2-sensitive genes (Regul), considering both up- and down- regulations.

(C) Global geometry of MCF-7 and MDA::ER E2-sensitive clusters defined using the criterion win=7 and Regul=3. These geometries were determined using a sliding window centered on each regulated genes included within the clusters (solid lines), or on each E2-regulated genes of the genomes (dashed lines). Distances are expressed as number of genes.

(D) The percentage of E2-regulated genes included within E2-sensitive clusters was calculated for each chromosomes.

(E) Genomic coordinates of the regions which were spotted on the arrays, which included clusters of E2-sensitive genes or single control genes. The No/Yes case for the chr14 region indicates that this region contains a single responsive gene in MCF-7 cells and a gene cluster in MDA::ER. The E2-sensitivity of the clusters in each cell line is indicated by ticks. The name of the MDA::ER top-regulated gene is indicated (Top), or the name of the MCF-7 top-regulated for MCF-7 specific clusters.

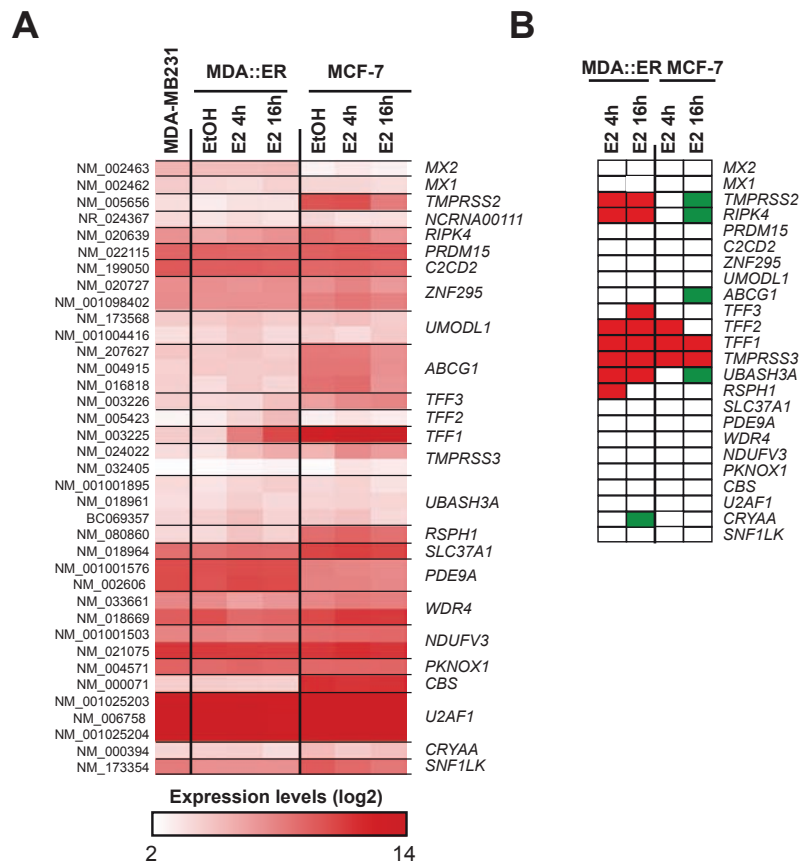


Figure S3. *Related to Figure 1.* Estrogenic regulations of genes included within the TFF cluster and neighborhood. **(A)** Heatmap of microarray expression values (log2) of RefSeq mRNAs from genes (depicted on the right side) included within the chr21 region of interest, in naive MDA-MB231, and MDA::ER or MCF-7 cells treated for 4h or 16h with 10^{-8} M Estradiol (E2) or Ethanol (EtOH) as vehicle control, as determined from the GSE23850. **(B)** Summary of the significant fold-changes (FC) of expression by E2, calculated from the GSE23850 values (FC>1.8 with a Benjamini-Hochberg corrected Student t-test p value <0.05. Activated genes are in red, whilst repressed ones are shown in green.

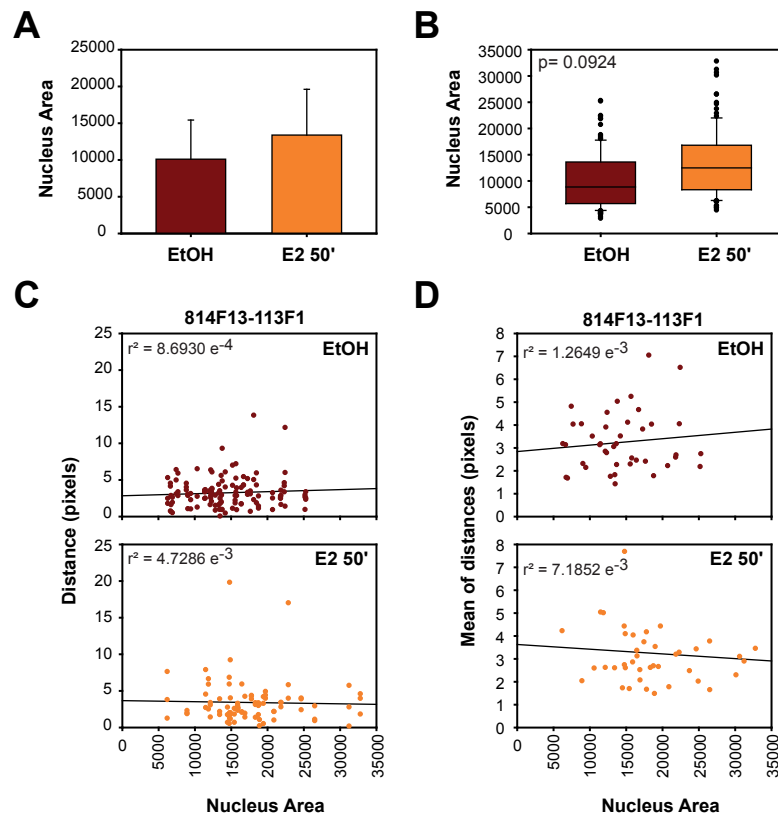


Figure S4. Related to Figure 2. Absence of correlation between nuclei area and paired probes distance.

(A and B) The nucleus area was calculated following the determination of a threshold fluorescence value corresponding to an entry transition into the nucleus. This value was manually determined as the inflection point of a profile plotting the DAPI signal measured in a 10 pixel large longitudinal window crossing the nucleus against pixel distance. These areas obtained in a representative FISH experiment (120 nuclei) are represented as mean values (A) or in boxplots (B) to examine their distribution upon E2 treatment. Fisher t-test indicates that the slight increase in the nucleus area is not significant.

(C and D) The distance between two paired signals (814F13 and 113F1 BAC probes) are represented as a function of the nucleus area. Individual distances of the three loci or their mean per nucleus are plotted in (C) and (D), respectively. These graphs clearly show that inter-probes distances and nuclei area are not correlated, in either control or E2-treated conditions.

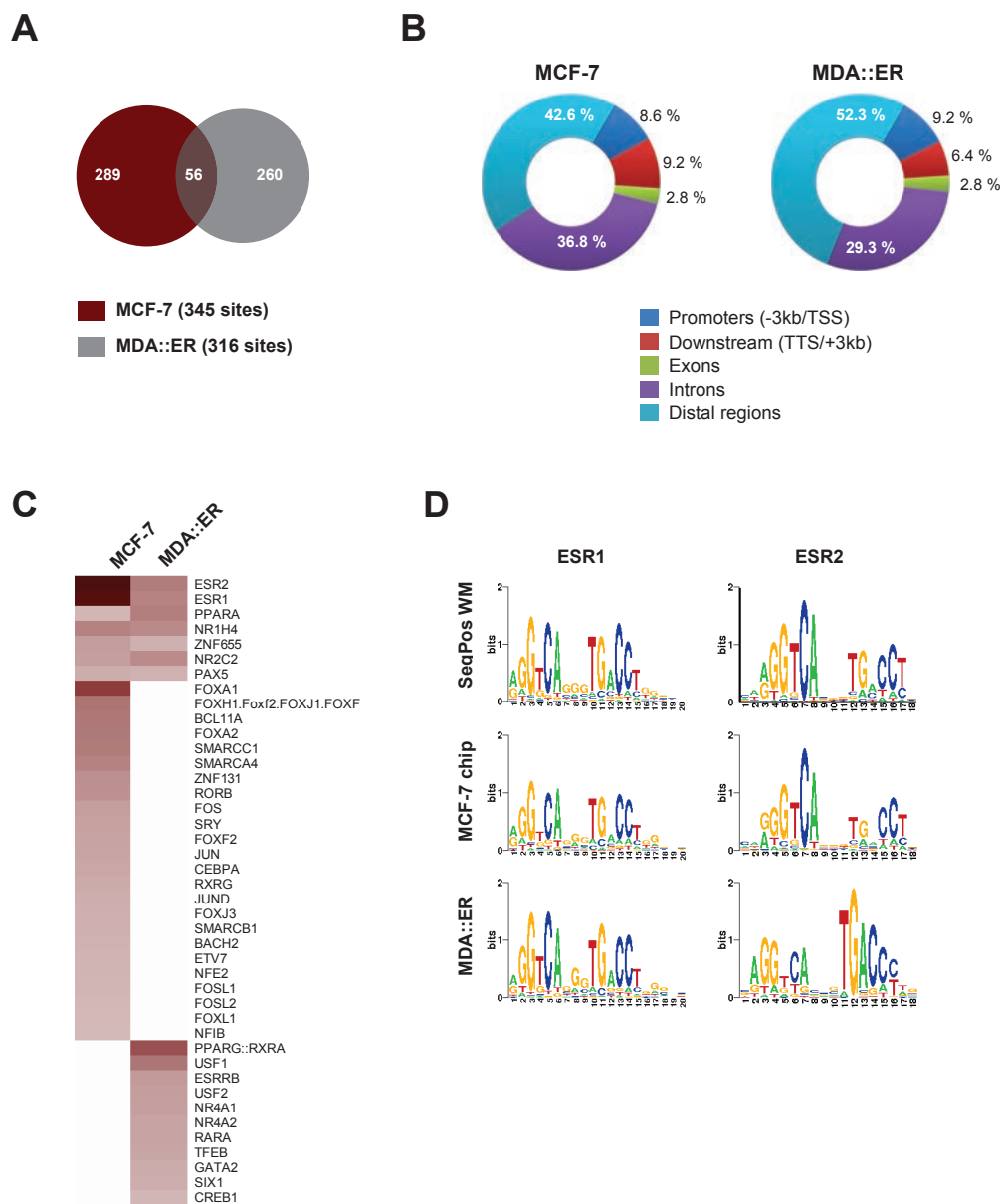


Figure S5. *Related to Figure 2.* ER recruitment sites in MDA::ER and MCF-7 cells within array-spotted genomic regions. (A) Overlap of ER binding sites (ER BS) determined in this study with those identified in MCF-7 cells by ChIP-chip (Carroll et al., 2006) on the genomic regions spotted on our cluster-specific arrays. The Venn diagram was generated under [cistrome](http://cistrome.dfci.harvard.edu/ap) (<http://cistrome.dfci.harvard.edu/ap>). (B) Genomic annotations of the distribution of ER BS towards specific gene features, as evaluated through the CEAS algorithm ([cistrome](http://cistrome.dfci.harvard.edu/ap)). (C) Enrichment of transcription factor motifs within the two populations of ER BS determined through ChIP-chip analysis (± 600 bp around the peak summit), as determined through the SeqPos algorithm ([cistrome](http://cistrome.dfci.harvard.edu/ap)). Results are shown as a clustered heatmap ranking calculated p-values for each motif. The significance of their enrichment was also filtered through their absolute Z-score value (threshold set at >2). Matrices that were also found enriched in sets of random regions (same number as ERBSs) were excluded from the heatmap. (D) ERE weight matrices defined for the ESR1 and ESR2 motifs identified in SeqPos.

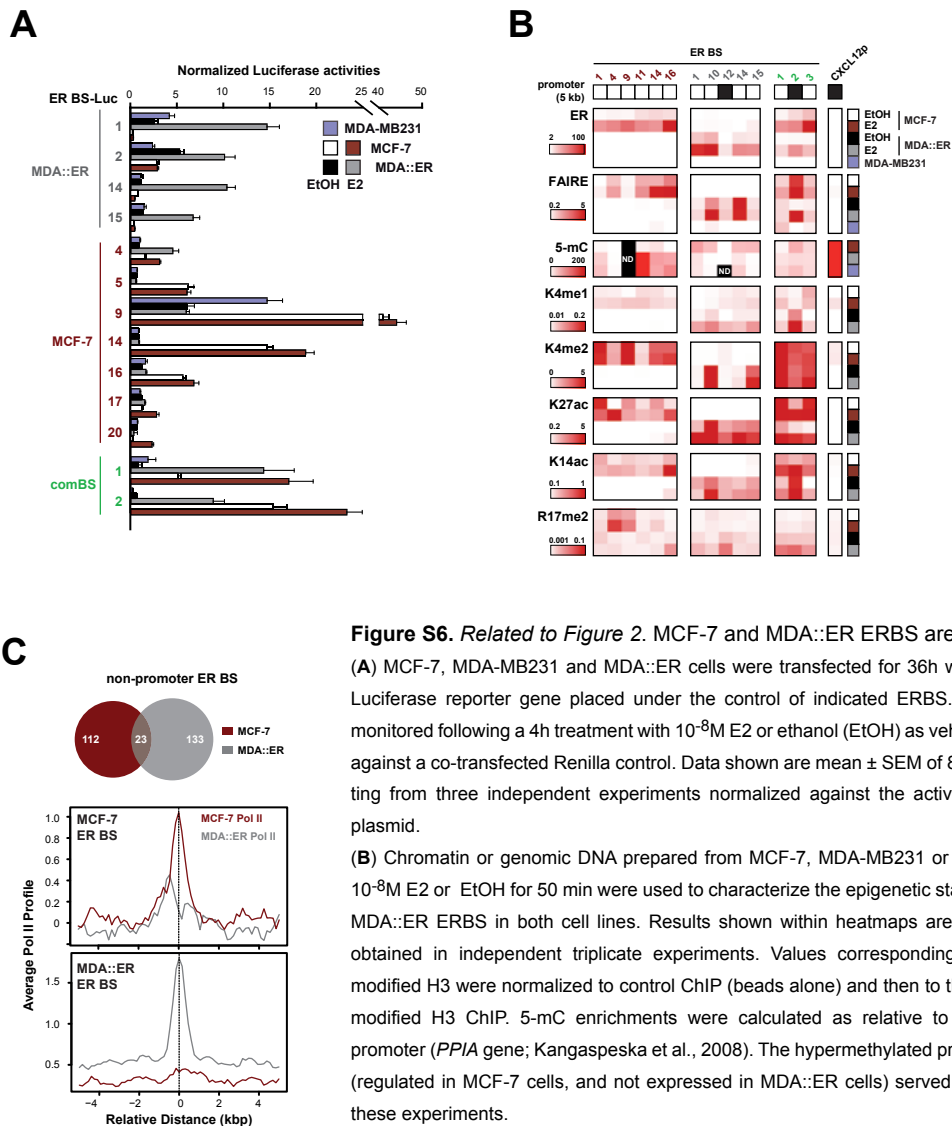


Figure S6. Related to Figure 2. MCF-7 and MDA::ER ERBS are cell-specific enhancers.

(A) MCF-7, MDA-MB231 and MDA::ER cells were transfected for 36h with plasmids containing the Luciferase reporter gene placed under the control of indicated ERBS. Luciferase activities were monitored following a 4h treatment with 10^{-8} M E2 or ethanol (EtOH) as vehicle control, and normalized against a co-transfected Renilla control. Data shown are mean \pm SEM of 8 to 9 sample values originating from three independent experiments normalized against the activity of the empty Luciferase plasmid.

(B) Chromatin or genomic DNA prepared from MCF-7, MDA-MB231 or MDA::ER cells treated with 10^{-8} M E2 or EtOH for 50 min were used to characterize the epigenetic status of indicated MCF-7 and MDA::ER ERBS in both cell lines. Results shown within heatmaps are means from 6 to 9 values obtained in independent triplicate experiments. Values corresponding to ChIP directed against modified H3 were normalized to control ChIP (beads alone) and then to those obtained in an anti-unmodified H3 ChIP. 5-mC enrichments were calculated as relative to those of an unmethylated promoter (*PPIA* gene; Kangaspeska et al., 2008). The hypermethylated promoter of the *CXCL12* gene (regulated in MCF-7 cells, and not expressed in MDA::ER cells) served as an additional control for these experiments.

(C) Enrichment signals obtained for anti-Polymerase II (Pol II) ChIP-chip experiments performed in cells-treated for 50 min with E2 (MCF-7 MAT data from Carroll et al., 2006) were aligned on MCF-7 or MDA::ER ERBS located more than 10 Kbp away from the TSS of any annotated gene. Data aggregation and plotting were done under Cistrome platform (SitePro tool). The overlap between these ER-dependent enhancers is illustrated by the Venn diagram at the top of the panel.

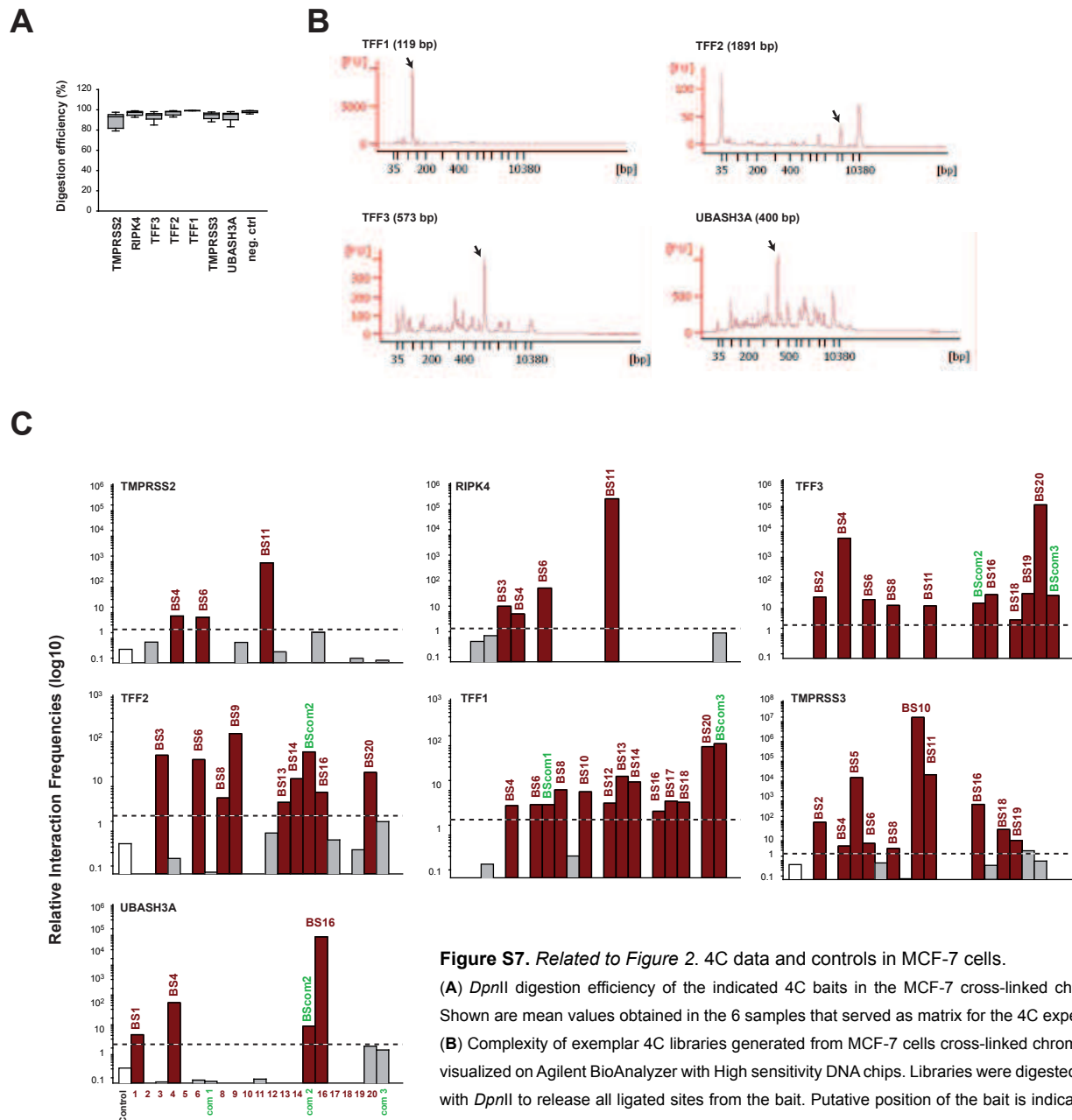


Figure S7. Related to Figure 2. 4C data and controls in MCF-7 cells.

(A) *DpnII* digestion efficiency of the indicated 4C baits in the MCF-7 cross-linked chromatin. Shown are mean values obtained in the 6 samples that served as matrix for the 4C experiments.

(B) Complexity of exemplar 4C libraries generated from MCF-7 cells cross-linked chromatin, as visualized on Agilent BioAnalyzer with High sensitivity DNA chips. Libraries were digested for 16h with *DpnII* to release all ligated sites from the bait. Putative position of the bait is indicated with arrows.

(C) A pool of six 4C libraries served as matrix for the specific amplification of target hybrid genomic regions. Values were normalized first to an internal positive control corresponding to the amplification of a DNA fragment originating from the ligation of the bait to its closest fragment, and then to internal negative control. Resulting values are plotted in the histogram, on a log10 scale. Interactions with a relative frequency below an arbitrary threshold of 2-fold over the negative control were considered as not significant (white bars). Positive interactions are indicated within each graphs in red with the name of the associated ERBS. First white bar (Control) corresponds to an additional negative region (*GAPDH* exon).

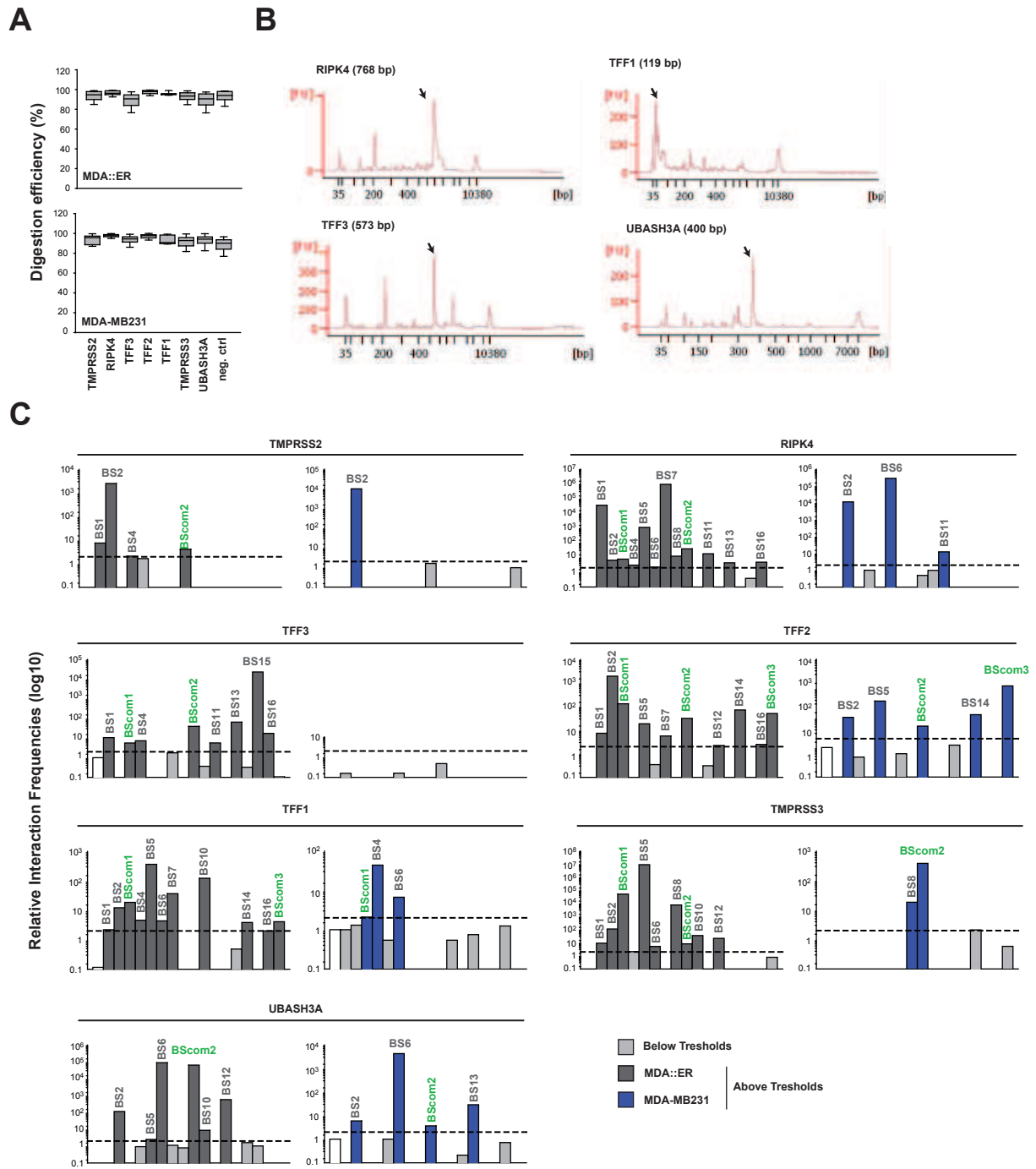


Figure S8. Related to Figure 2. 4C data and controls in MDA::ER and MDA-MB231 cells.

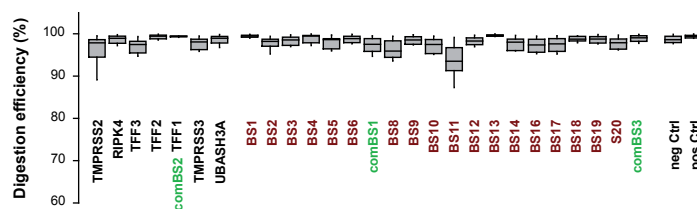
(A) *DpnII* digestion efficiency of the indicated 4C baits in MDA::ER and MDA-MB231 cross-linked chromatin. Shown are mean values obtained in the 6 samples that served as matrix for the 4C experiments.

(B) Complexity of some of the 4C libraries generated from MDA::ER cells cross-linked chromatin, as visualized on NanoDrop High sensitivity DNA chips. Libraries were digested for 16h with *DpnII* to release all ligated sites from the bait.

(C) A pool of six 4C libraries served as matrix for the specific amplification of target hybrid genomic regions. Values were normalized first to an internal positive control corresponding to the amplification of a DNA fragment originating from the ligation of the bait to its closest fragment, and then to internal negative control. Resulting values are plotted in the histogram, on a log10 scale. Interactions with a relative frequency below an arbitrary threshold of 2-fold over the negative control were considered as not significant (white bars). Positive interactions are indicated within each graphs with the name of the associated ERBS. First white bar (Control) corresponds to an additional negative region (*GAPDH* exon).

A

RESULTATS-PARTIE I



B

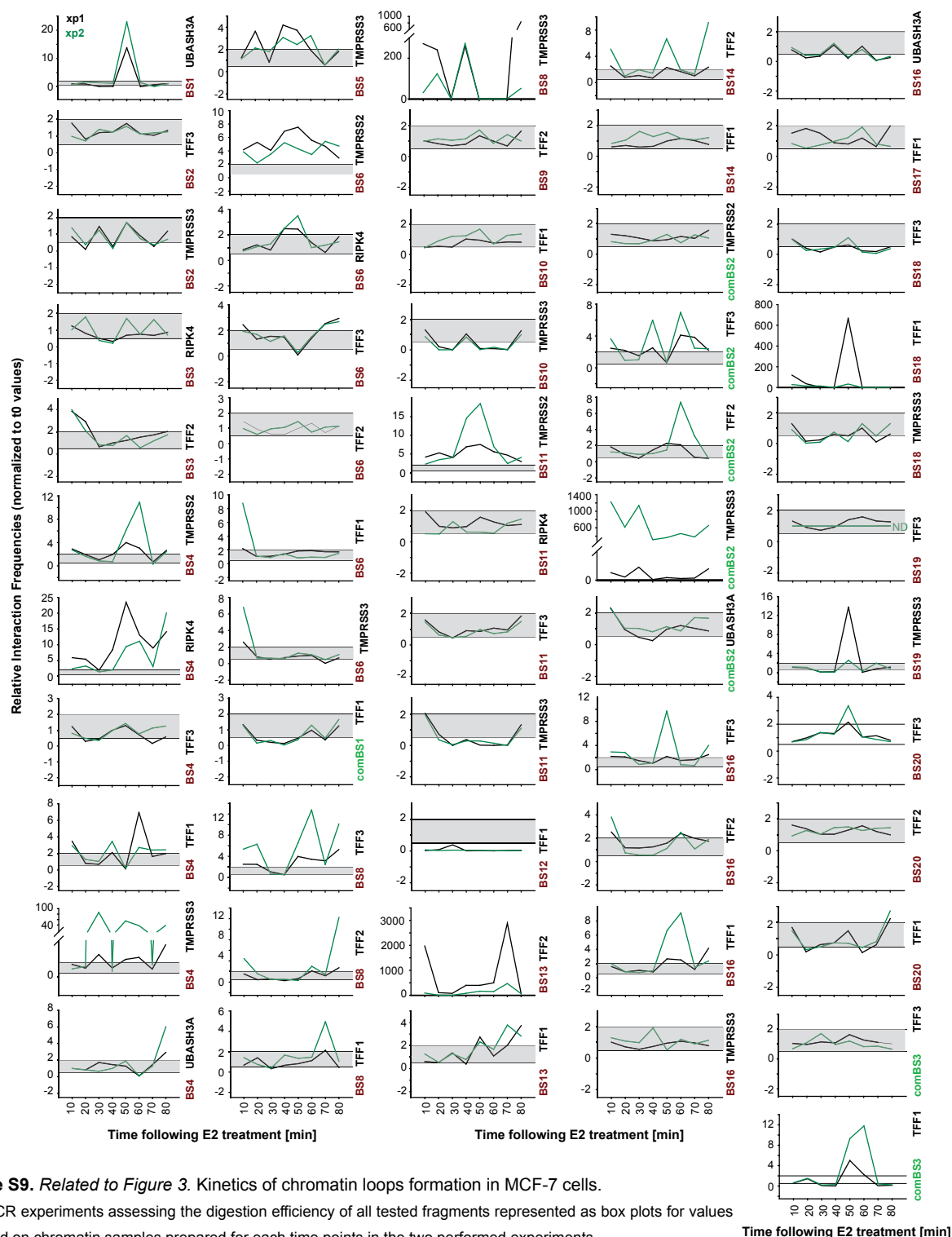


Figure S9. Related to Figure 3. Kinetics of chromatin loops formation in MCF-7 cells.

(A) qPCR experiments assessing the digestion efficiency of all tested fragments represented as box plots for values obtained on chromatin samples prepared for each time points in the two performed experiments.

(B) 3C-qPCR experiments testing the relative interaction frequencies between ERBSs and genes promoters indicated on the right of each sub-panel. Values obtained in the two different experiments are shown (xp1 and xp2), and were calculated as indicated in the supplementary materials and methods section. The grey box in each graph depicts the significance threshold applied to these analysis: variations were declared significant when varying from control (t0) from at least 2-fold.

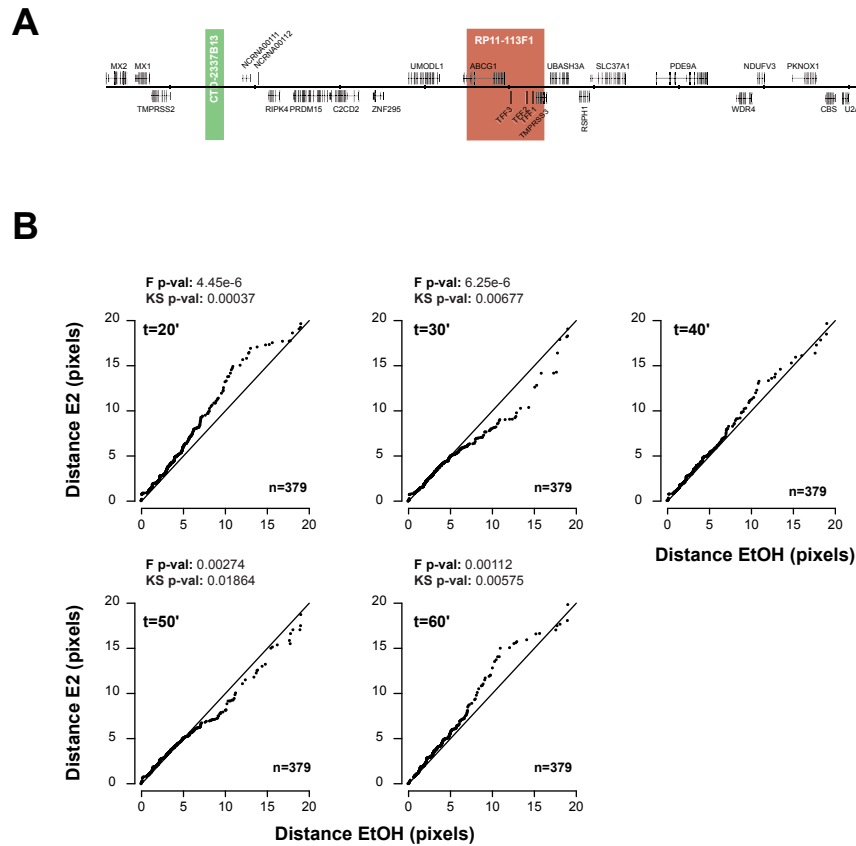


Figure S10. Related to Figure 3. Kinetic DNA-FISH.

(A) Positions of the BAC used to generate fluorescent probes.

(B) DNA-FISH was performed on fixed cells at indicated times following the addition of E2 at 10^{-8} M final concentration. The distribution of distances separating paired FISH probes are represented within quantile-quantile (Q-Q) plots representation of conditionally ranked measured distances for each time of treatment vs. the basal condition (ethanol, EtOH). The normal distribution expected from non-varying distances is illustrated by the straight black line in each Q-Q plots. Non-parametric Fisher test and Kolmogorov-Smirnov p-values (F and KS, respectively) are indicated when determined as significant.

[illegible]

Figure S11. Related to Figure 4. BLASTs controls for TFOs. Summary of BLAST searches for TFOs target sequences and TFOs on *homo sapiens* genome, DNA and miRNA databases (BLAST, <http://blast.ncbi.nlm.nih.gov/>; miRbase, <http://www.mirbase.org/>). Only top5 hits are shown (top5 genes for the blastn RNA) when possible. Significance threshold for miRNA screening was set to an e-value of 50.0. Sequences are indicated, with nucleotides in red highlighting the oligopyrimidine-oligopurine tracts targeted by the TFOs. Mismatches from strict polyA/G sequences are in italic. Locations of the TFOs are indicated as well (d), expressed in bp according to the ERBS peak center.

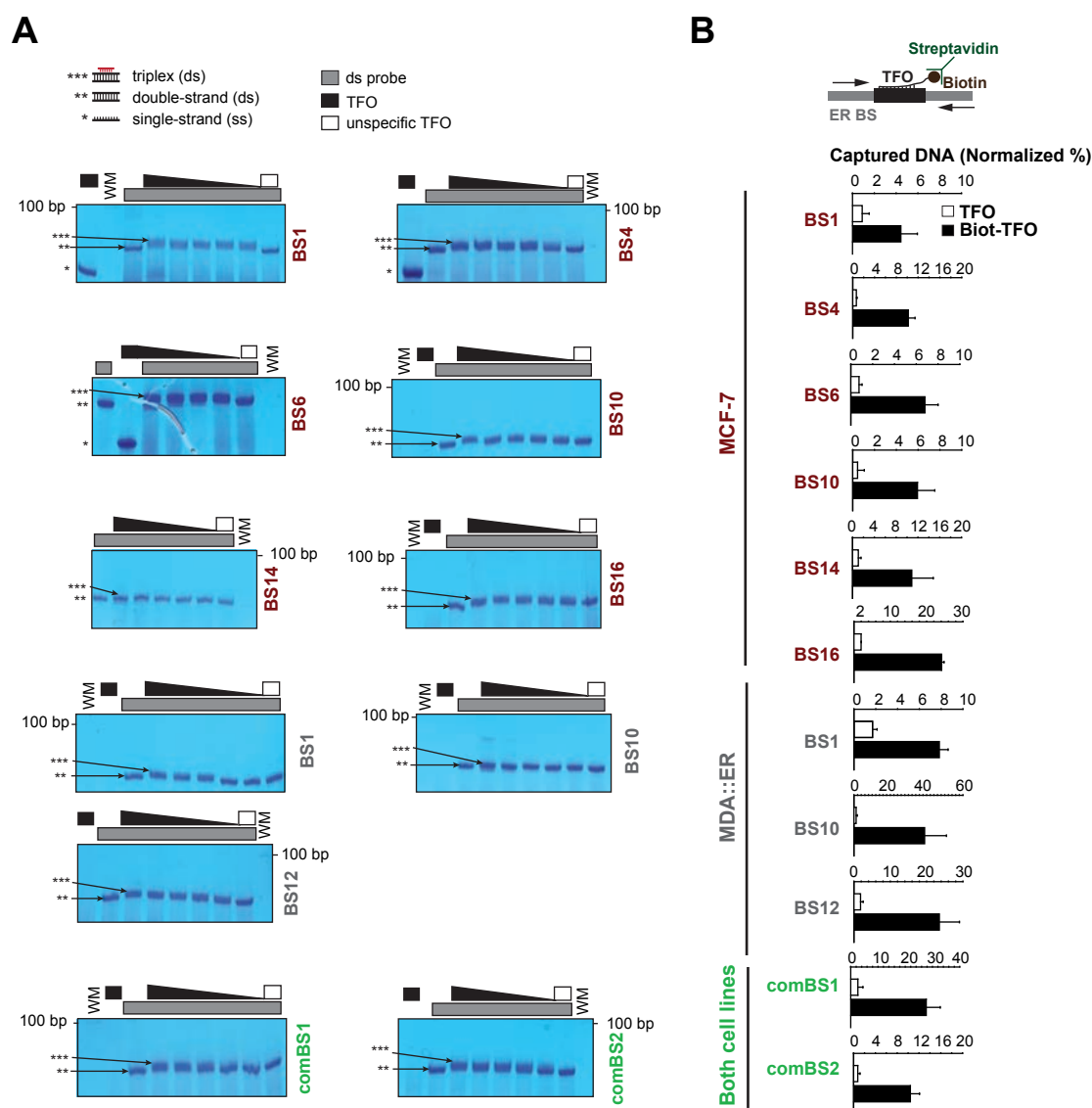
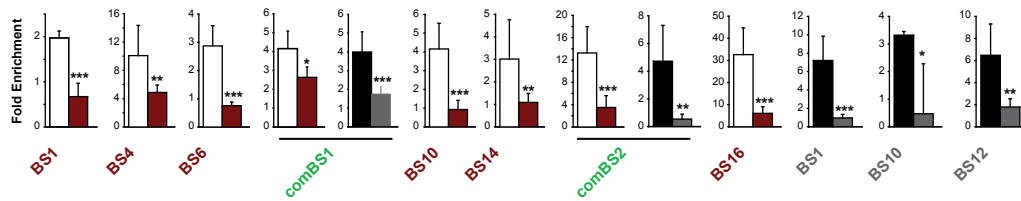


Figure S12. Related to Figure 4. *In vitro* monitoring of TFO-mediated DNA triplex formation and target recognition.

(A) Formation of DNA triplex was analyzed by gel-shift. Increasing amounts of TFOs (25 pmol to 1.5 nmol) were added to 25 (MCF7 ERBS 1 and ERBS 4) or 50 pmoles of DNA duplexes (sequences in Supplementary Figure 19) and incubated for 16h at 37°C in a buffer containing 10 mM MgCl₂, 100 mM NaCl, 50 mM Tris-HCl (pH 7.4), 10% glycerol and 0.5 mg/ml tRNA. TFOs at the highest concentration or duplexes alone were processed as other samples in order to provide internal controls for these experiments. An additional control was performed by incubating duplexes with an unspecific TFO at the highest concentration. Complexes were separated by electrophoresis on 15% (19:1) polyacrylamide gel containing 10 mM MgCl₂ and 50 mM Tris-HCl (pH 7.4) at room temperature, and visualized by methylene blue staining. WM lane was for standard weight markers.

(B) Triplex capture assays. MCF-7 and/or MDA::ER cells were transfected for 36 h with 10 μmol of TFOs or biotinylated (Biot-) TFOs. Following cross-linking, cells were lysed and sonicated chromatin was then incubated with streptavidin-coated beads. Amounts of captured DNA were analyzed by qPCR. Values shown are mean ± SD of two independent duplicates, and are expressed as % of captured DNA relative to input DNA (ΔCt) normalized to the amounts of recovered negative control region (*Rplp0* promoter).

A



B

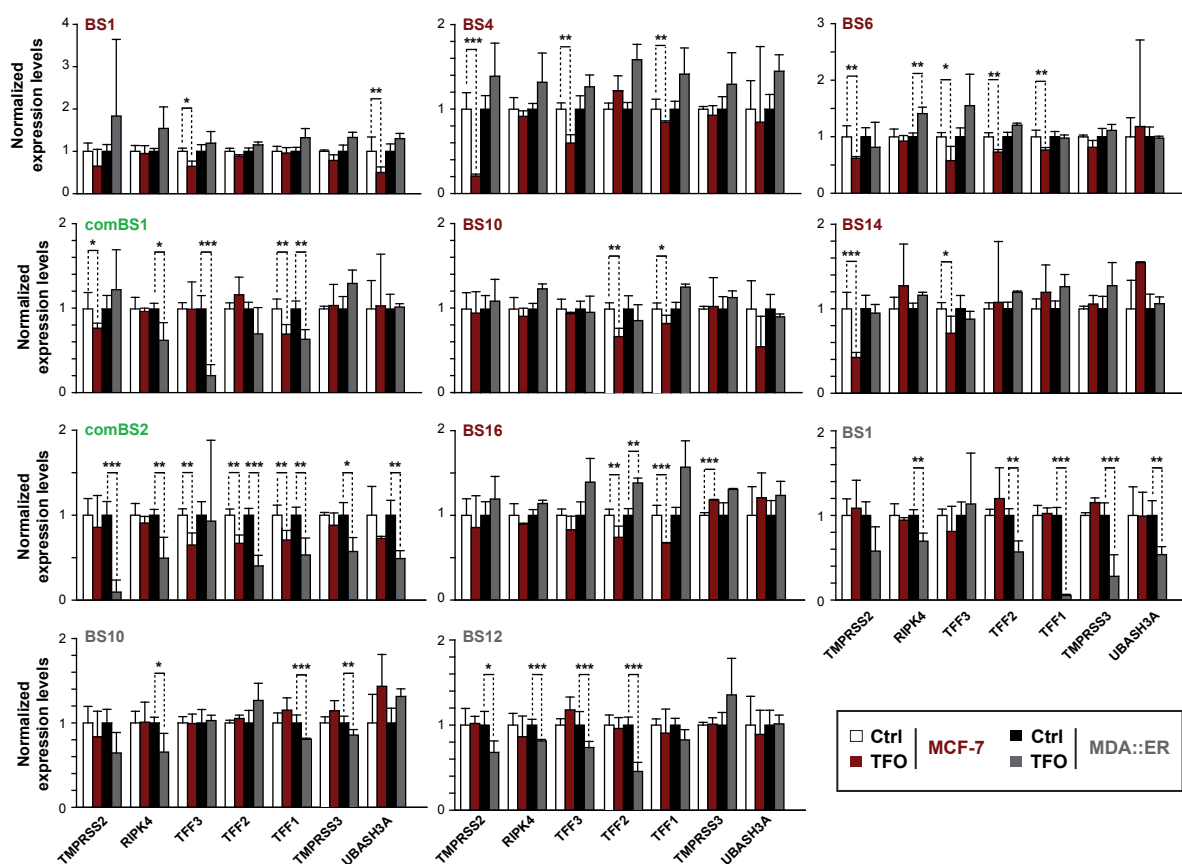


Figure S13. Related to Figure 4. Impact of TFOs on ER binding and gene expression.

(A) MCF-7 and MDA::ER cells were transfected for 36 h before a 4 h treatment with 10^{-8} M E2 with 10 μ mol of TFOs. Anti-ER ChIP-qPCR were then performed. Data shown are normalized to those obtained in a control ChIP (beads alone) and then to the normalized values obtained using oligonucleotides amplifying a control sequence (*PKNOX* promoter).

(B) Following transfections performed as in A, cells were scraped and lysed for RNA preparation and RT-qPCR experiments. Expression values for each of the tested genes were normalized to those of a control gene (*PKNOX*) and then to normalized values from a control sample transfected with a control TFO. Data shown are mean \pm SD of 8 to 11 samples originating from three independent experiments. Wilcoxon test was used to identify statistically relevant variations: p-value <0.001 (***), <0.01 (**), <0.05 (*).

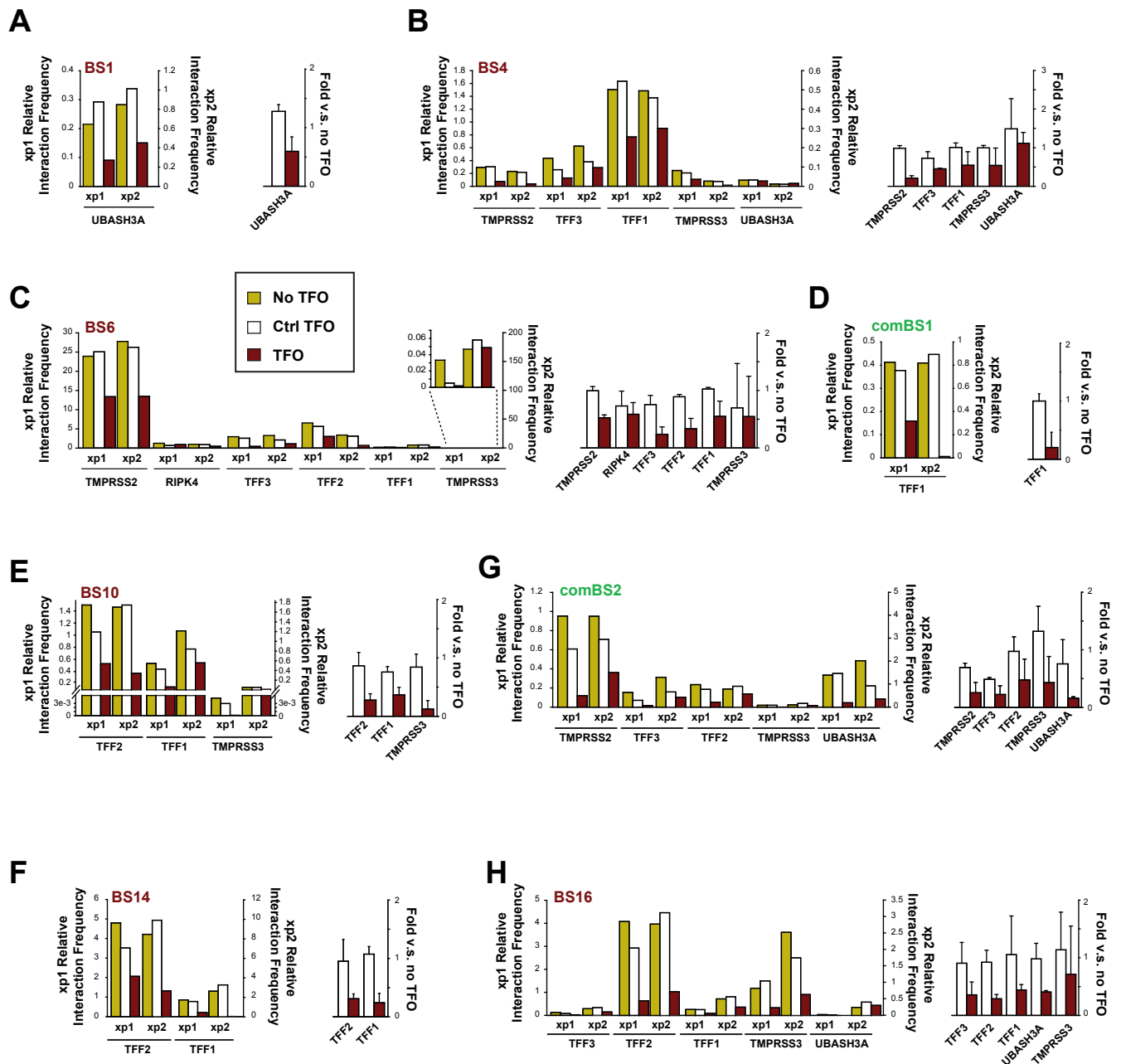


Figure S14. *Related to Figure 4.* Impact of TFOs on the three-dimensional organization of the TFF cluster in MCF-7 cells.

(A to H) 3C experiments were conducted in MCF-7 cells following a 36 h transfection with 10 μ M TFOs and 50' treatment with 10⁻⁸M E2. Results of both independently performed experiments (xp) are depicted within each subpanels, as well as the mean fold-changes of interaction frequencies provoked by the transfection of either control (Ctrl) or specific TFO as compared to untransfected cells. As all along the manuscript, data were normalized to internal controls and digestion efficiency, and expressed as relative to an estrogen-insensitive chromatin loop. Effects of TFOs on the formation of chromatin loops were considered significant only when control TFO had no impact on the interaction and when the absolute fold-changes were above 0.5 log units.

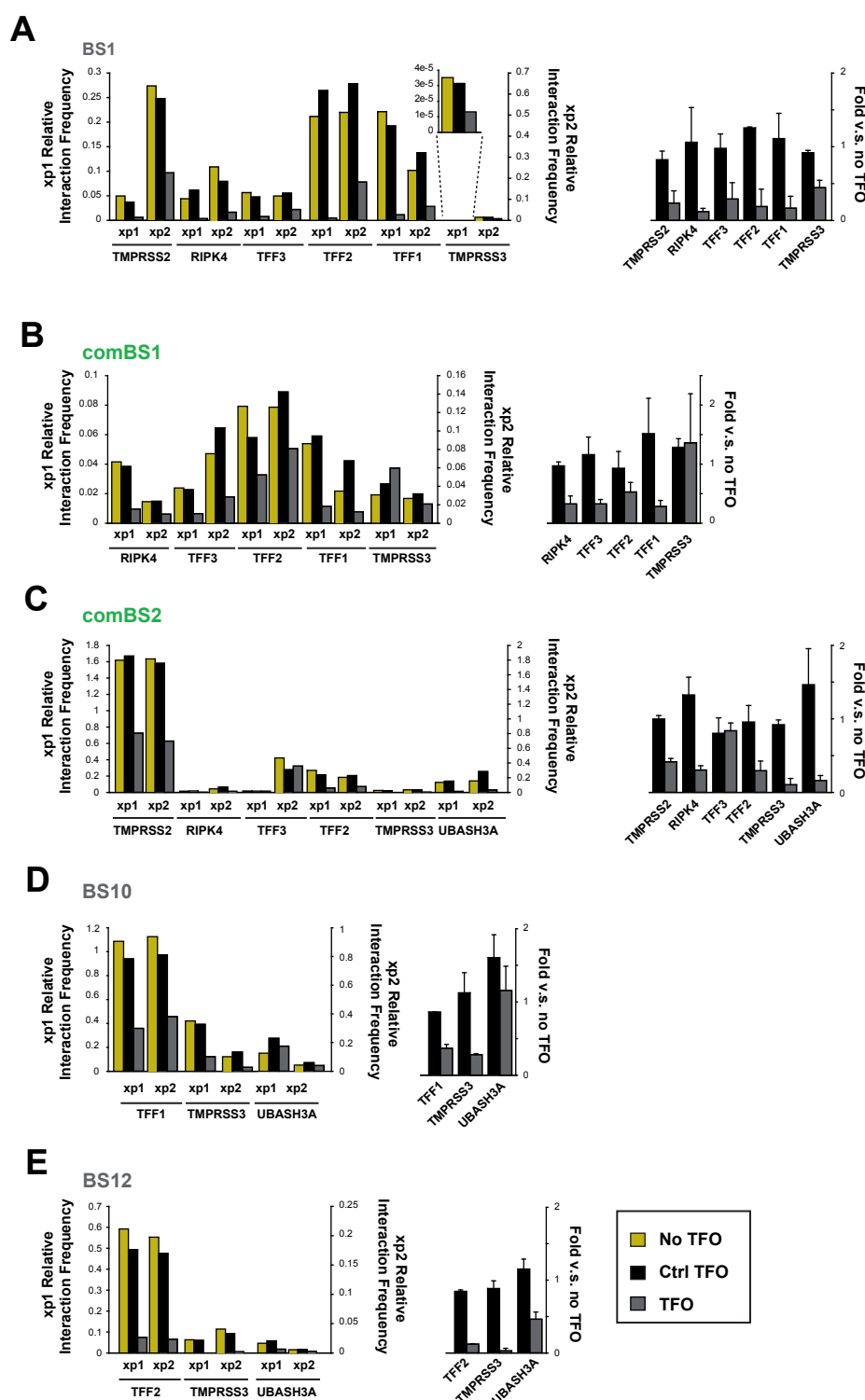
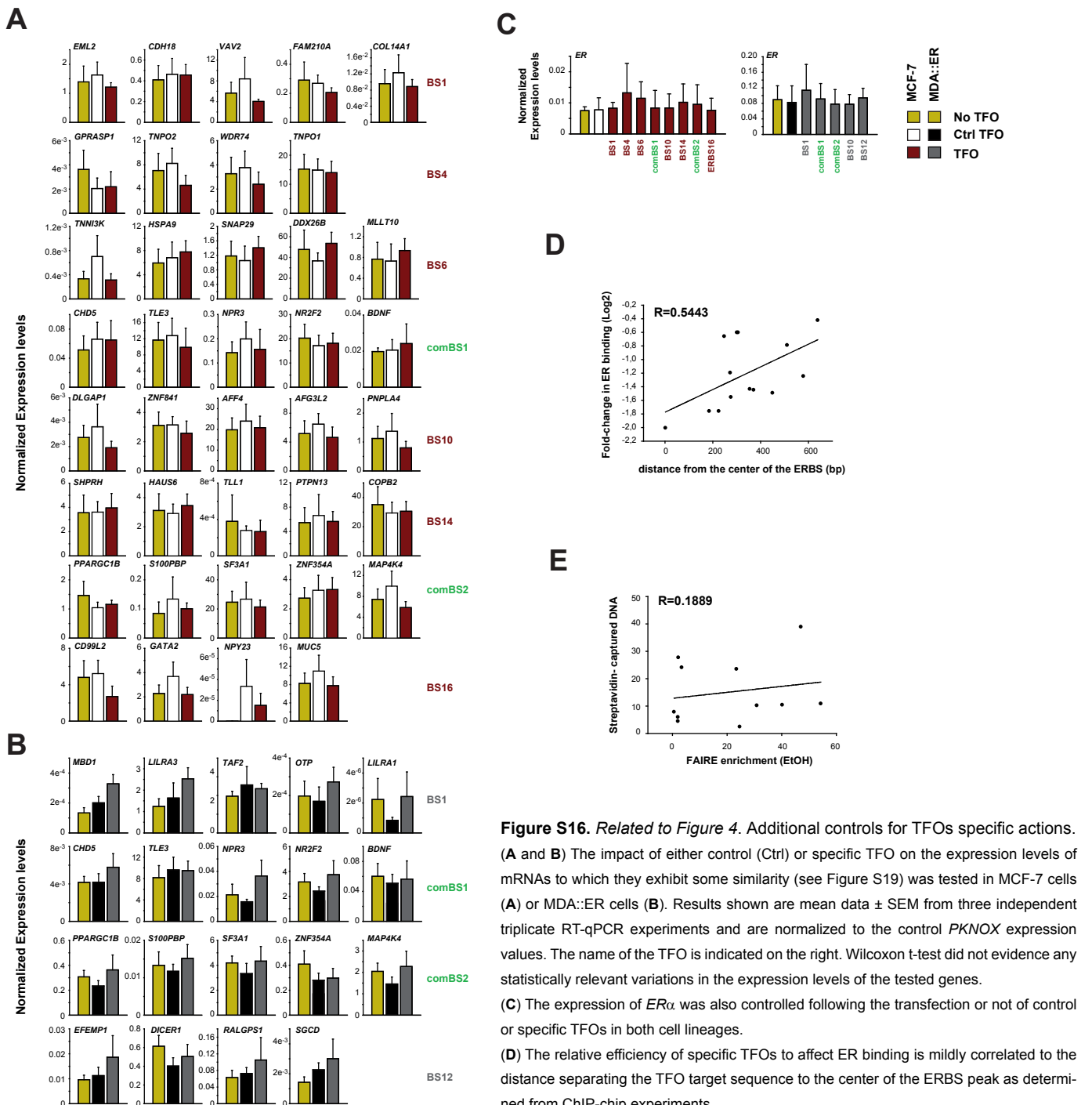


Figure S15. *Related to Figure 4.* Impact of TFOs on the three-dimensional organization of the TFF cluster in MDA::ER cells. (A to E) 3C experiments were conducted in MDA::ER cells following a 36 h transfection with 10 μ M TFOs and 50' treatment with 10⁻⁸M E2. Results of both independently performed experiments (xp) are depicted within each subpanels, as well as the mean fold-changes of interaction frequencies provoked by the transfection of either control (Ctrl) or specific TFO as compared to untransfected cells. As all along the manuscript, data were normalized to internal controls and digestion efficiency, and expressed as relative to an estrogen-insensitive chromatin loop. Effects of TFOs on the formation of chromatin loops were considered significant only when control TFO had no impact on the interaction and when the absolute fold-changes were above 0.5 log units.



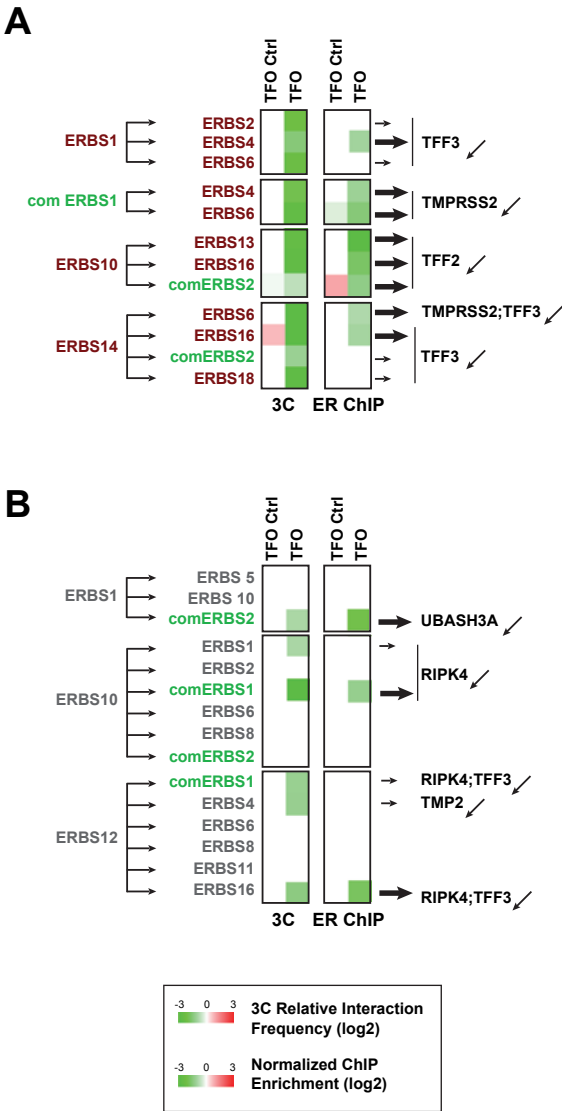


Figure S17. Related to Figure 4. TFOs also perturbate ERBS-ERBS functional interactions.
(A and B) The impact of either control (Ctrl) or specific TFO on the interactions made between indicated ERBS (identified by independent 3C experiments, not shown) was tested by 3C-qPCR and ER-ChIP experiments in MCF-7 cells (A) or MDA::ER cells (B). 3C results shown are those of a representative experiment from two and are expressed as log2 relative interaction frequencies normalized to those measured in untransfected cells. ER ChIP-data summarized in the heatmaps are mean values significantly differing from the ones obtained in untransfected cells, expressed as log2 and obtained in 2 independent triplicate experiments. On the right side of the figure is represented the transcriptional collateral effect that could be associated with the identified TFO mediated impact on ERBS-ERBS interaction and on ER mobilization on distant ERBS.

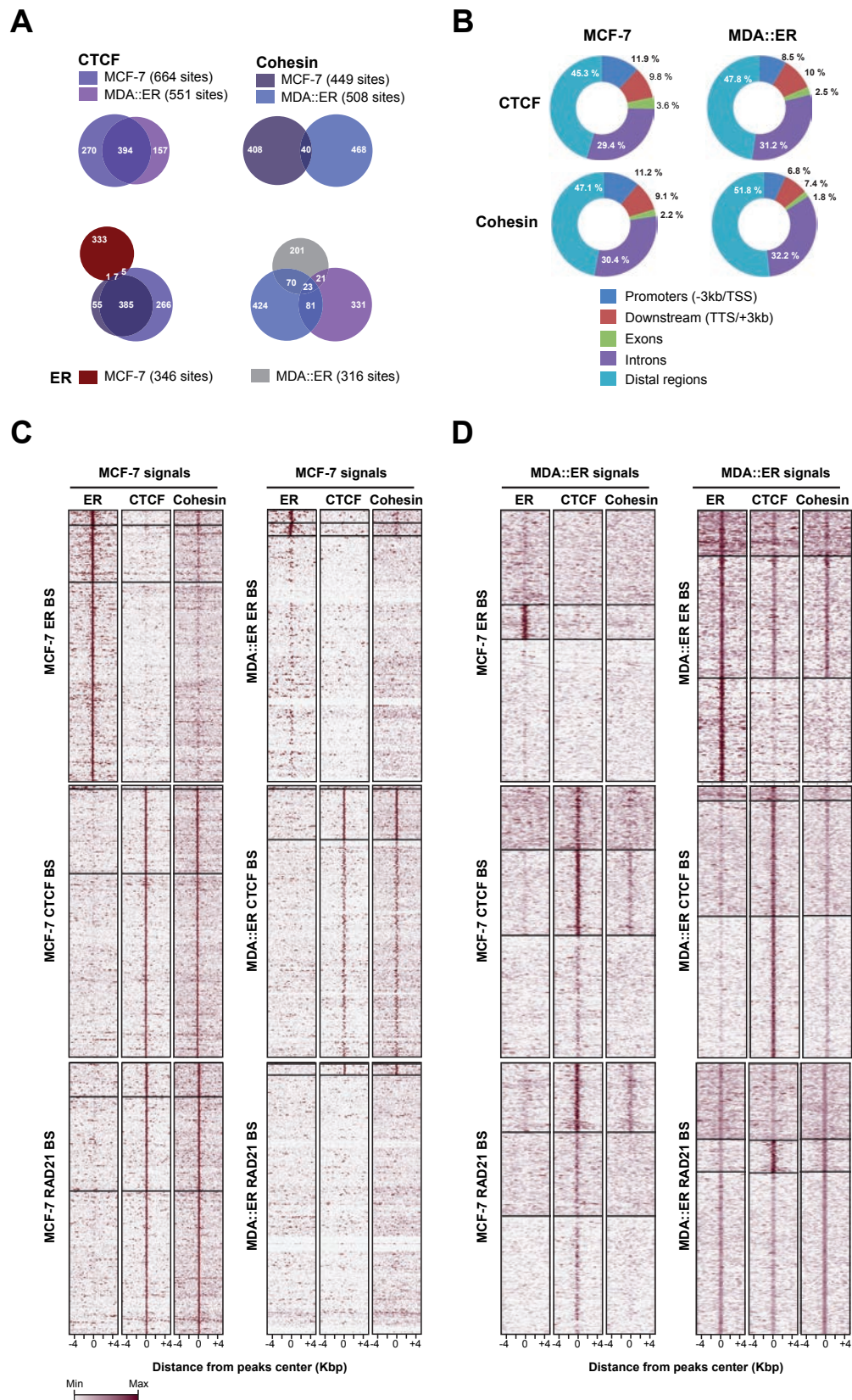


Figure S18. Related to Figure 5. CTCF and cohesin recruitment sites in MDA::ER and MCF-7 cells within array-spotted genomic regions. (A) Overlap of CTCF and cohesin binding sites determined in this study (p-value<0.005) with those identified in MCF-7 cells (Schmidt et al., 2010 ; GSE25021) on the genomic regions spotted on our cluster-specific arrays. Venn diagrams were generated under cistrome website. (B) Genomic annotations of the distribution of CTCF and cohesin BS towards specific gene features (CEAS). (C and D) Heatmaps generated under the cistrome website of MCF7-ChIP-Seq (C) or MDA::ER ChIP (D) signals on aligned MCF-7 or MDA::ER genomic regions (+ surrounding 5 kbp of the peaks center) mobilizing either ER, CTCF or the cohesin complex. These results, indicate that the overlap between ER, CTCF and cohesin binding sites might be more extended than the Venn diagrams tend to show.

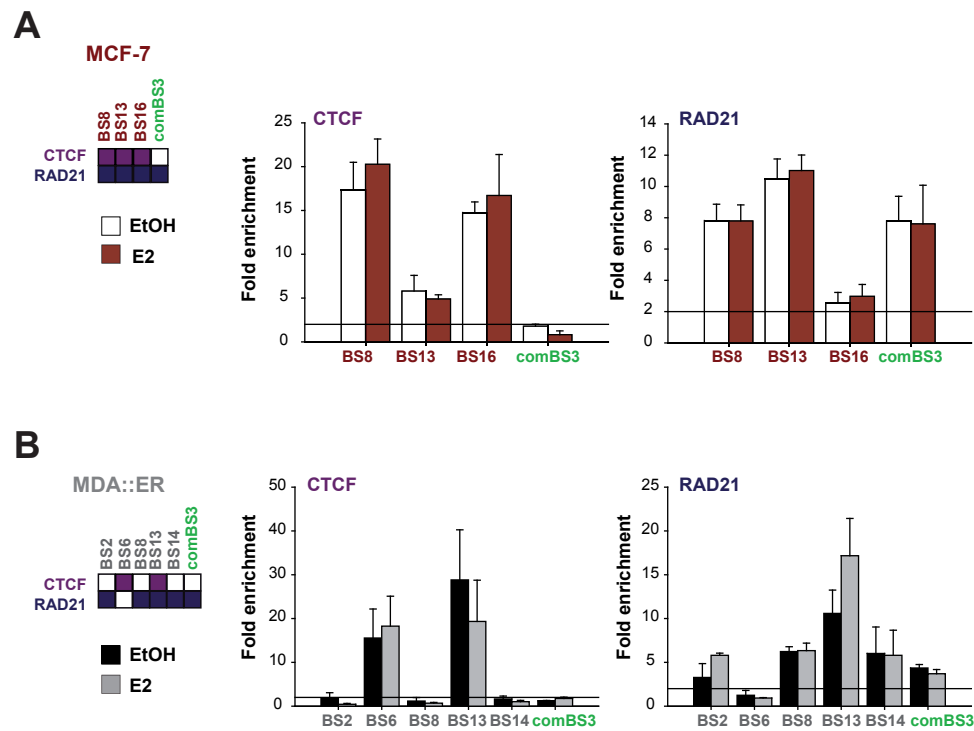


Figure S19. Related to Figure 5. CTCF and RAD21 recruitments are not affected by E2.

CTCF and RAD21 ChIP were performed on chromatin prepared from MCF-7 (**A**) or MDA::ER (**B**) cells treated for 4h with 10^{-8} M E2 or ethanol (EtOH) as vehicle control. Fold enrichment of the precipitated proteins on tested sequences was normalized over control ChIP and negative region (*PKNOX1* promoter) values. Data shown are mean values \pm SD obtained in three independent triplicate experiments.

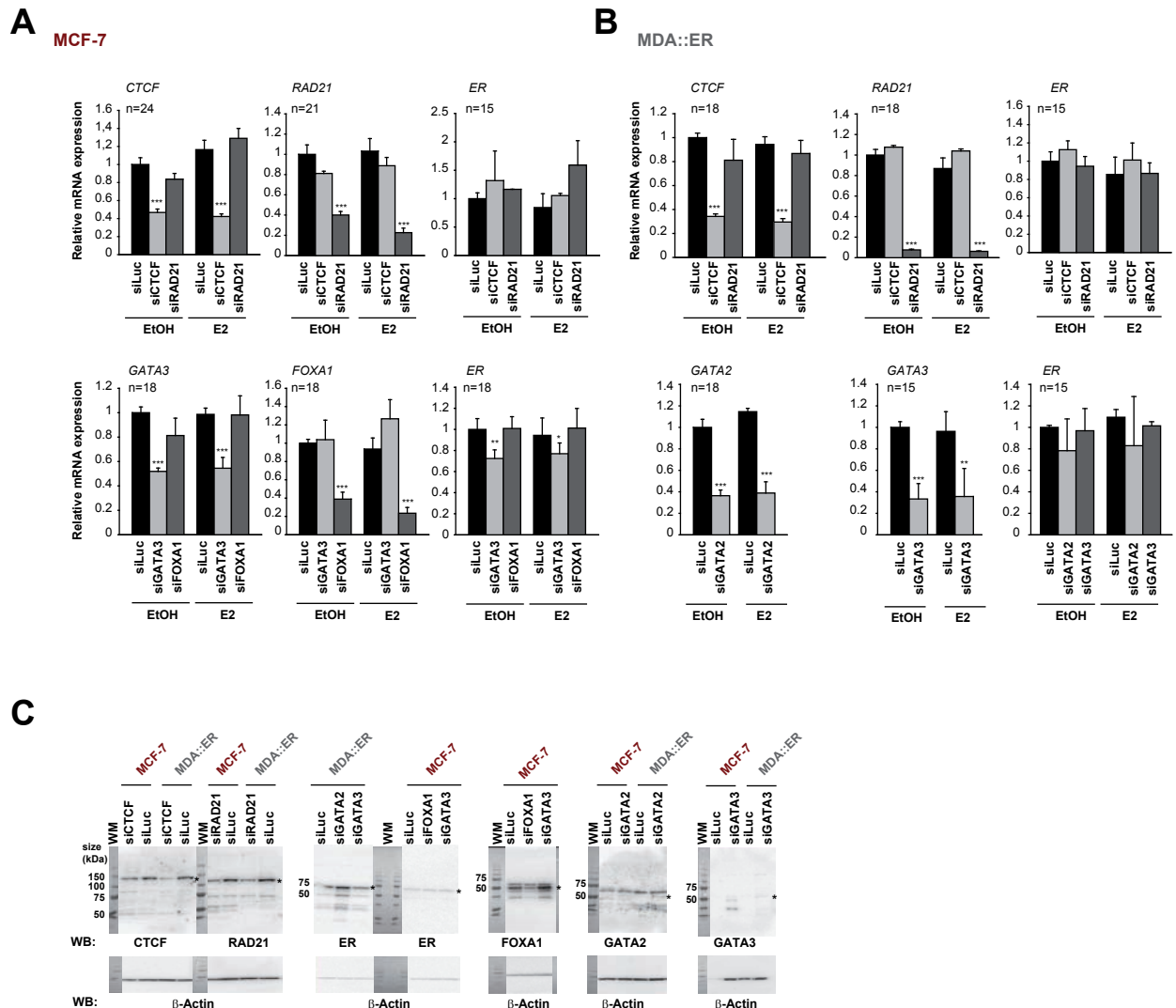


Figure S20. siRNAs controls.

(A and B) RT-qPCR were performed to assess the efficiency of the siRNAs used in this study. Indicated mRNA amounts were evaluated in MCF-7 and MDA::ER cells treated with E2 for 4h or ethanol (EtOH) as vehicle control following 72h of transfection with indicated siRNAs. Results shown are mean data \pm SEM of values obtained in at least 5 different experiments, normalized to the expression of the control *PKNOX* gene. Wilcoxon test was used to identify statistically relevant variations from control siLuc/EtOH condition: p-value <0.001 (***), <0.01 (**).

(C) Representative Western blots (WB) for protein samples prepared from cells transfected with indicated siRNAs. The relative amounts of deposited material were checked by probing for β -Actin. The bands corresponding to the proteins of interest are denoted with an asterisk. WM stands for weight markers.

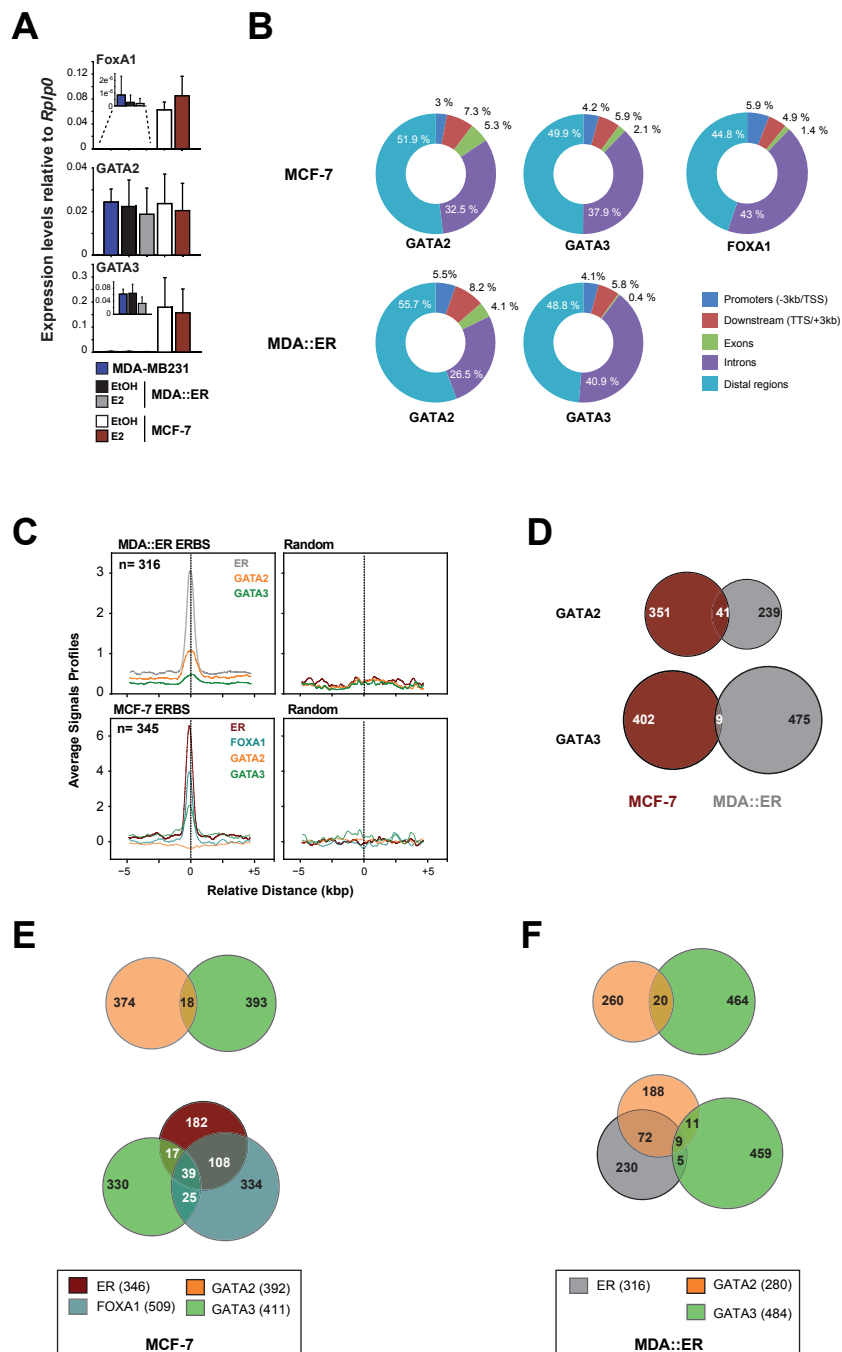


Figure S21. Related to Figure 6. Recruitment of FOXA1 and GATA proteins in MDA::ER and MCF-7 cells .

(A) RT-qPCR detecting indicated mRNA amounts in MDA-MB231 and MDA::ER or MCF-7 cells treated with E2 for 4h or ethanol (EtOH) as vehicle control.

(B) Genomic annotations of the distribution of FOXA1 and GATAs binding sites towards specific gene features, as evaluated through the CEAS algorithm (cistrome).

(C) The ER, GATA1, GATA2, GATA3 and FOXA1 ChIP-chip signals (MAT scores for MCF-7 ER and FOXA1 experiments, MA2C scores for the others) were averaged and aligned using SeqPos (cistrome) on the ERBS regions determined in each cells, or on the same number of random regions exhibiting similar mean length (right side of the panel).

(D to F) Venn diagrams indicating the overall overlaps of ER, FOXA1, GATA1, GATA2 and GATA3 cistromes in MCF-7 and MDA::ER cells. BS for each GATA protein are compared between both cell line in D; the overlap between GATA1, GATA2 and GATA3 BS in each cell at the top of panels E and F; and the global overlaps of GATA3/FOXA1 or GATA1/GATA2/GATA3 with MCF-7 or MDA::ER ERBS in the bottom part of panels E and F, respectively.

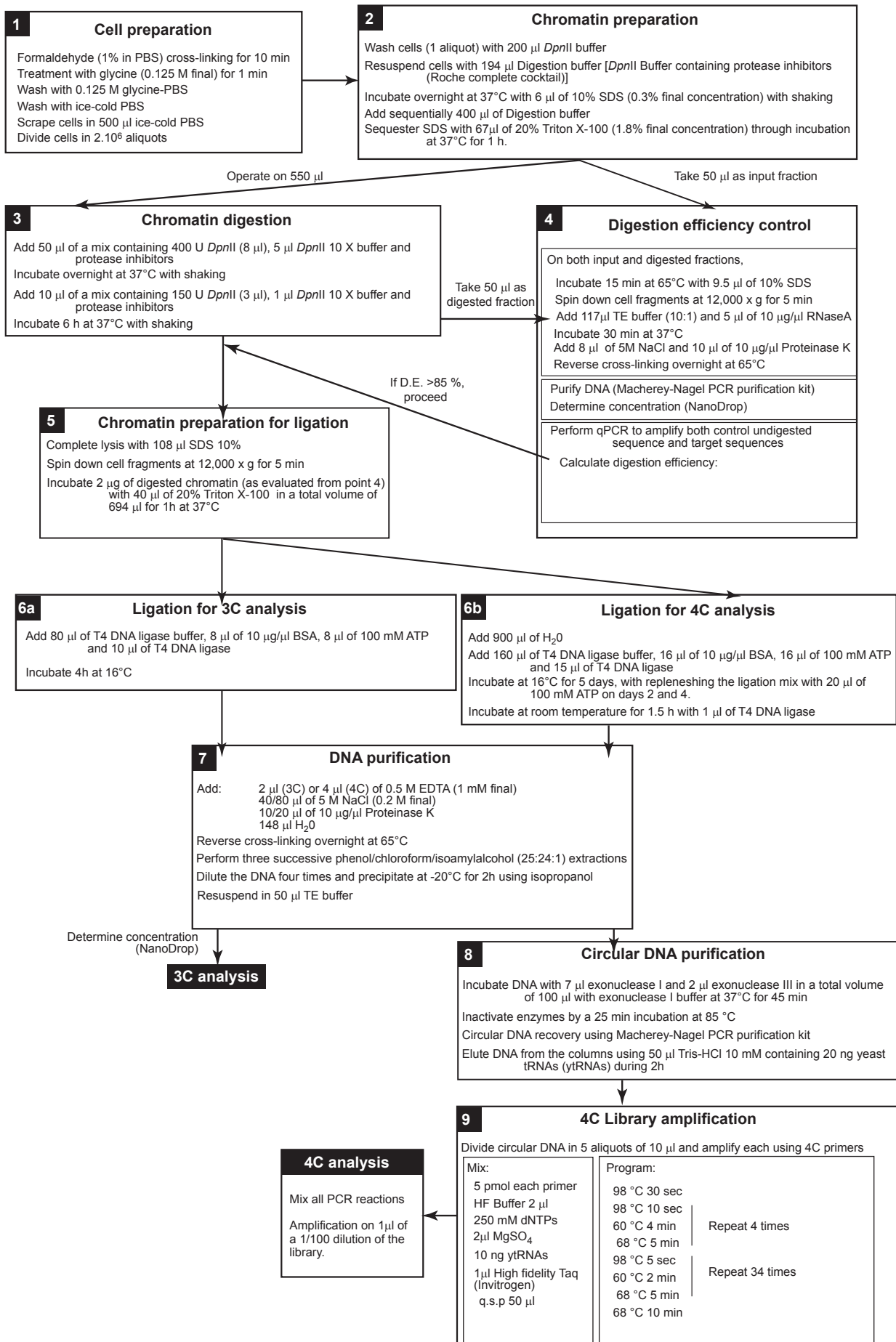


Figure S22. Support for Experimental Procedures. Workflows of 3C and 4C procedures.

RÉSULTATS | CHAPITRE II

CONTEXTE DE L'ETUDE

La première partie des travaux a porté sur l'identification des mécanismes à l'origine de la régulation d'un cluster de gènes E2-dépendant, le cluster TFF, dans deux types cellulaires différents, les MCF-7 qui expriment ER de façon endogène et les MDA::ER qui expriment ce récepteur de façon ectopique. Nous avons pu montrer que la régulation de ce cluster mettait en jeu des mécanismes semblables mais des partenaires protéiques différents. Nous avons également mis en avant une différence importante dans les transcriptomes œstrogéno-dépendants et dans les cistromes d'ER entre ces deux lignées cellulaires. Des études faites entre autre dans le laboratoire de M. Brown (Carroll *et al.*, 2006, Hurtado *et al.*, 2011) ont montré le rôle primordial de FOXA1 dans la régulation E2-dépendante et la mise en place des sites de recrutement de ER. FOXA1 est un facteur pionnier dont l'action est de fonctionnaliser la chromatine pour la liaison d'autres facteurs, ici ER, et cela à travers la di-méthylation de la lysine 4 de l'histone H3. Cette protéine est présente dans les MCF-7 mais elle n'est pas exprimée dans les MDA::ER. Nous avons donc émis l'hypothèse que les partenaires protéiques différents et plus particulièrement l'absence de FOXA1 dans les MDA::ER pouvait être à l'origine des différences fonctionnelles observées entre les MCF-7 et dans les MDA::ER.

OBJECTIFS

Le but de cette étude était donc de déterminer si l'introduction de FOXA1 dans les MDA::ER permettait d'orienter le cistrome ER et le transcriptome E2-dépendant de ce nouveau type cellulaire (MDA::ER::FOXA1) vers des profils de type MCF-7, mais également si le phénotype prolifératif des MCF-7 en présence d'E2 était restauré.

CONTRIBUTIONS AUX DIFFERENTS RESULTATS

- Sélection et de caractérisation des clones (expériences de Western blot, RT-qPCR)
- Préparations des ADNc pour les analyses transcriptomiques
- Réalisation des ChIP ER (MDA::ER::pCDNA et MDA::ER::FOXA1) en vue de leur séquençage, ainsi qu'une partie des ChIP H3K4me2, H3K27ac et FOXA1 qui ont été séquencés. Réalisation des ChIP ER et FOXA2 validant les données de séquençage sur une vingtaine de sites aléatoires
- Expériences de Me-DIP et de hMe-DIP
- Evaluations par RT-qPCR des expressions de FOXA1, ER dans les différents types cellulaires utilisés et RT-qPCR de validation, MDA::ER et les MCF-7
- Réalisation des expériences de ChIP ER, FOXA2, H3K4me2 et H3K27ac et de RT-qPCR sur les gènes du cluster TFF après déplétion de FOXA2 par siRNAs.

A FOXA1/FOXA2 competition may hinder the concerted reorientation of an ER negative cell line towards an estrogen-sensitive growth phenotype by FOXA1 and ER

Justine Quintin, Christine Le Péron, Elise Mahé, Gaëlle Palierne, Maud Bizot, Frédéric Percevault, Stéphane Avner, Gilles Salbert, and Raphaël Métivier.

Equipe SP@RTE, UMR CNRS 6290. Equipe labellisée Ligue contre le Cancer Université de Rennes I. Campus de Beaulieu. 35042 RENNES CEDEX. FRANCE.

Contact:

Raphaël Métivier, Equipe SP@RTE, UMR CNRS 6290. Batiment 13. Université de Rennes 1. Campus de Beaulieu. 35042 Rennes Cedex. Tel: +33 (0)2 2323 5142 Fax: +33 (0)2 2323 6794; E-mail: raphael.metivier@univ-rennes1.fr

SUMMARY

The primordial actions of the estrogen receptor alpha (ER) in controlling breast cancer cells proliferation rate are particularly documented. However, reintroducing ER into ER-negative cells does not reprogram their transcriptome towards an estrogen-sensitive growth phenotype. Genome wide studies together with functional explorations have shown that a member of the Forkhead (FKH) family of transcription factors, FOXA1, is essential for the appropriate binding of ER onto chromatin in MCF-7 cells. FOXA1 pioneering actions involve modulations of histone post-translational modifications (HPTMs) and local depletion in methylated CpGs. Using MDA-MB231 cells which are ER and FOXA1 negative, we aimed at determining whether the introduction of ER and FOXA1 would be sufficient to restore an estrogen-sensitive growth and to coincidentally reprogram chromatin. We used ChIP-seq experiments to map ER and FOXA1 binding sites and to profile H3K4me2 and H3K27ac marks, and surprisingly observed that FOXA1 had rather a global negative impact on ER actions, associated with an inhibition of growth and a reorientation of the transcriptome of the cells towards cell death. We demonstrate that the re-introduction of FOXA1 in these cells actually competes for the binding and pioneering actions of another FKH protein, FOXA2. As demonstrated for its *alter ego* FOXA1 in MCF-7 cells, abrogating FOXA2 expression drastically diminished ER binding onto its cognate sites, associated with a local loss of HPTMs for active chromatin. It also prevented the occurrence of all mechanisms required for the estrogenic regulations of the entire *TFF* genomic loci. Finally, silencing FOXA2 seems to lift the mechanistic lever that blocks FOXA1 influence on ER binding, thereby suggesting that its presence in MDA::ER cells may hinder the influence of ectopically expressed FOXA1 on cell proliferation.

INTRODUCTION

Breast cancer is the most common hormone-dependent cancer in woman worldwide (Kanavos, 2006). The growth of the majority of these cancers is directed by the 17β -estradiol (E2), whose actions are mainly mediated by the estrogen receptor α (ER). Direct physiological roles of this protein are also known in other systems such as reproduction, bone, cardiovascular biology (Russo and Russo, 2006). ER behaves as a ligand-dependant transcription factor (TF), and belongs to the nuclear receptor superfamily of TFs (McEwan, 2009). As other TFs, the main regulatory mechanism mediated by ER is to bind to specific genomic regions either directly to specific recognition sequences termed EREs (Estrogen Responsive Elements) or indirectly through protein/protein interactions. Common interactors for these secondary elements are AP1 or Sp1 TFs for instance (Hall et al., 2001; Nott et al., 2009). Once tethered onto DNA, it is now known for long that ER can recruit a variety of cofactors or corepressors to regulate the transcription of its nearby target gene (Shang et al., 2002; Metivier et al., 2003). During the last decade, genome-wide studies have however challenged this view of ER-directed gene regulations (Eeckhoutte et al., 2009; Farnham, 2009; Magnani et al., 2011a). First, the recruitment of ER occurs in the majority of the cases on sequences called enhancers, located far away from any annotated E2-sensitive gene (Carroll et al., 2005; Carroll et al., 2006; Wellboren et al., 2009). Hence, the characterization of the chromatin loops linking such distant elements to their target gene, as recently achieved using the ChiA-PET technology (Fullwood et al., 2009), has been a major challenge in order to fully apprehend the three-dimensional network that organizes an E2-sensitive transcriptome.

Second, these genome-wide studies have also demonstrated that the actions of other TFs are essential for ER binding accuracy and frequency. FOXA1 is one of the most studied of these proteins (Carroll et al., 2006; Hurtado et al., 2011), termed pioneer factors since their action would be to prepare chromatin for other TF binding, presumably through acting as recruitment platform or signal for histone modification and/or nucleosome remodeling complexes (Cirillo et al., 1998; Cirillo and Zareth, 1999; Cirillo et al. 2002). FOXA1 actions on chromatin have been shown to be also associated with methylation and/or hydroxymethylation of DNA (Serandour et al., 2011; 2012). PBX1 (Magnani et al., 2011b) or AP2 (Tan et al., 2011) are other proteins that may act as pioneer factors for ER mobilization onto chromatin.

We previously demonstrated that the expression of ER in MDA-MB231 cells (MDA::ER clonal cell line) generates an E2-sensitive transcriptome that shares very little overlap with that of an E2-dependant cell line such as MCF-7 cells (Legagneux *et al.*, in preparation; Quintin *et al.*, submitted). Moreover, as also observed by others, this forced expression marginally reduced the proliferation of the cells (Garcia *et al.*, 1992). Here, we explored the hypothesis that the absence of FOXA1 in MDA::ER cells may account for these differing responses to E2. Using ChIP-seq as main approach, we were able to show that the ectopic expression of FOXA1 in this particular cell system does not restore a MCF-7 like ER cistrome because of a pioneering action of FOXA2 onto ER binding.

MATERIAL AND METHODS

Chemicals, Enzymes, Antibodies. All chemicals, restriction or modification enzymes were obtained from Sigma, Roche or New England Biolabs. Antibodies (**Table S1**) were purchased from Abcam, Diagenode, Millipore, NeoMarkers or SantaCruz.

Cell culture and Establishment of Stable Clones. MCF-7, MDA-MB231 and MDA-MB231 cells stably expressing ER α (MDA::ER; Reid et al., 2003) were maintained in DMEM (Sigma) supplemented with 10% fetal calf serum (FCS, Sigma) and antibiotics (Roche) at 37°C under 5% CO₂. MDA::ER cells were grown in DMEM/5% antibiotics and Hygromycin (Calbiochem, 0.8 mg/ml). To generate FOXA1-expressing MDA::ER cells (MDA::ER::FOXA1) and their associated control cells (MDA::ER::pCDNA), MDA::ER cells were first stably transfected with the pcDNA6/TR vector (Invitrogen). Clones were selected using 5 μ g/ml Blastidicin (Invitrogen) and screened for tetracycline inducibility following the transient transfection of the reporter pcDNA4/TO/Luc construct (Invitrogen) and monitoring of Luciferase activities. A positive clone with the lowest measured Luciferase background (reflective of a maximal repressing activity of the Tet repressor) was next transfected with a pcDNA4/TO/FOXA1 construct (Serandour et al., 2011) or the empty control plasmid. Selection of stably transfected cells was performed using a combination of 5 μ g/ml blastidicin and 75 μ g/ml Zeocin (Invitrogen) and Hygromycin. Clones were screened for FOXA1 expression using RT-qPCR and Western blot assays on RNAs or cellular extracts prepared from cells grown for 2 days in the presence of 1.5 μ g/ml tetracycline (Sigma). For experiments that required cells to be treated with E2, cells were cultivated for 3 days in 9cm \emptyset dishes in DMEM/2.5% charcoal-stripped FCS (csFCS) and required selective antibiotics prior to any treatment.

FACS analysis. Following trypsinisation, 1.5 10^6 cells were collected in PBS containing 30% immunofunctional assay (IFA) buffer (10 mM HEPES (pH 7.4), 150 mM NaCl, 4% FCS) and fixed with 5 ml of cold ethanol (EtOH) 70% for 30 min on ice. Cells were then washed twice in PBS and placed at -20°C in 3 ml EtOH 70 % overnight. Following centrifugation at 200 x g for 6 min, the cells were washed again with PBS and then permeabilized in 1 ml of IFA buffer containing 0.5 % of Tween-20. Following recovery by centrifugation, cells were incubated for 15 min at 37°C in 500 μ l of PBS complemented with RNase A (10 μ g/ml final concentration). Single cell suspension was then obtained by filtration through nylon meshes, and incubated with propidium iodide (40 μ g/ml final concentration) for 1h in obscurity before analysis on a FACScan apparatus (Becton Dickinson).

Reverse-Transcription. Total RNAs from 10^7 cells were purified using TrizolTM reagent (Life Technologies, Inc.) according to the manufacturer's instructions. Two μ g of RNA served as template for M-MLV reverse transcriptase (Invitrogen) and Pd(N)6 random hexamers (Amersham Pharmacia Biosciences).

Expression arrays. RNAs for microarray analysis were isolated from 2.10^7 MDA::ER::pCDNA or MDA::ER::FOXA1 cells treated for 48h with 1.5 μ g/ml tetracycline and with 10^{-8} M E2 or ethanol (EtOH) as vehicle control using the RNeasy Plus Mini Kit (Qiagen) with homogenization through QIAshredders (Qiagen). Integrity and purity of the RNAs were controlled on an Agilent Bioanalyzer using the RNA 6000 Nano Assay (Agilent). Ten μ g of selected samples exhibiting a RIN >9.8 were then subjected to cDNA synthesis using the Superscript Double-Stranded cDNA Synthesis Kit (Invitrogen) and a mix of 50 pmol random hexamers and 50 pmol of Oligo-dT.

cDNAs were then treated for 10 min at 37°C with 5 µg RNaseA (Invitrogen), purified through a phenol:chloroform:isoamyl alcohol extraction on Phase Lock light gels (Fisher Scientific) and then precipitated. Following quality controls (migration of 250 ng cDNA on agarose/BET gels and qPCR on control genes), all subsequent steps (labeling of 2.5µg cDNA, hybridization and scanning of the arrays) were performed at the NimbleGen service facilities (Reykjavic, Island). For each experimental condition, three arrays (NimbleGen Homo sapiens 385K Array) were hybridized with independently prepared pools of cDNAs synthesized from experimental triplicates (independent experimental and biological triplicates). Quantile normalization of the data through the RMA algorithm and all primary analyses were performed using the ArrayStar software suite (DNASTar, Inc.). Regulated genes were identified using home-made R scripts to calculate paired fold-changes (FC) in gene expression, and were considered significant for FC > 1.8 and with a p-value < 0.05 as determined by a Benjamini-Hochberg corrected Student t-test. Interrogations for enrichments in GO terms and biological pathways were made under the Webgestalt platform (www.bioinfo.vanderbilt.edu/webgestalt/; Duncan et al., 2010).

Chromosome conformation capture (3C). The 3C method, adapted from Gondor et al. (2008) and Hagege et al. (2007) was previously described (Quintin et al., submitted), and used the *DpnII* enzyme.

DNA-IP, FAIRE and ChIP assays. Me-DIP protocol, adapted from Mohn et al., 2009, used 4 µg of fragmented DNA and 2 µg of specific antibody, as described in Serandour et al., 2011. Formaldehyde-assisted isolation of regulatory elements (FAIRE) methodology and ChIP experiments were conducted as previously reported (Serandour et al., 2011 and Quintin et al., submitted, respectively).

ChIP-Seq experiments. ChIP experiments for NGS analysis were performed following the same procedures than those previously described in Quintin et al. (submitted). Ten to fourteen ChIP samples prepared in four to 5 independent experiments were purified, pooled and concentrated through ethanol precipitation. Construction of libraries and sequencing using Illumina Genome Analyzer II were conducted at the IGBMC sequencing facility (Strasbourg, France). Reads were aligned onto indexed chromosomes of genome hg19 using bowtie (Langmead et al., 2009) with parameters allowing at most one mismatch and selected for unique mapping onto the genome. Sequencing statistics are given in **Table S2**. Extracted reads were then converted to .wig signal files, using samtools (Li et al., 2009) and MACS (Zhang et al., 2008) with default parameters. To minimize the bias of diverging sequencing depths between different ChIP samples prepared with the same antibody, the signal intensities of a given .wig were linearly normalized so as to be comparable to the .wig file with the highest sequencing depth. Peak callings were then performed using the algorithm described in Serandour et al., 2012 with a *p-value* for the threshold set to $1e^{-8}$ and peaks defined as being constituted of at least 4 adjacent signals within 65 bp above the threshold values which are given in **Table S2**.

Bioinformatics and statistics. Genomic annotations were performed under the CEAS algorithm present at the cistrome web-platform (<http://cistrome.dfci.harvard.edu/ap/>; Blankenberg et al., 2007) and under the GREAT web-platform (<http://bejerano.stanford.edu/great/public/html/index.php>; McLean et al., 2010). Searches for enriched motifs for TF binding were executed under cistrome using the SeqPos tool (He et al., 2010) or at the CentDist (Zhang et al., 2011) webpage

(http://compbio.ddns.comp.nus.edu.sg/~chipseq/webseqtools/TASKS/Motif_Enrichment/submit.php).

Heatmaps of qPCR results were generated under Mev (Saeed et al., 2003) and represent mean values that were declared as significantly different from controls by Wilcoxon or Student paired t-tests. To facilitate the reading of the heatmaps, the mean value of an experimental sample was replaced by the control one when it did not significantly vary.

Quantitative PCR (qPCR). All qPCRs were performed on BioRad CFX96 machines using BioRad iQ SYBR Green supermix. Primers for RT-PCR were designed under QuantPrime (<http://www.quantprime.de/>; Arvidsson et al., 2008). Oligonucleotides for all other type of samples were designed under Primer3 (<http://frodo.wi.mit.edu/primer3/>). Sequences for all primers (Sigma) are available upon request.

RESULTS

FOXA1 does not reprogram the transcriptome of cells ectopically expressing ER towards a generic cell response to E2. To explore the hypothesis that the pioneer FOXA1 transcription factor might be the key lacking factor in MDA::ER cells to recapitulate an estrogeno-sensitive cell growth as it is observed in MCF-7 cells, we first tried to generate an MDA::ER cell lineage stably expressing FOXA1. However, we were not able to isolate a single clone of such cells. The extremely limited number of surviving cells exhibited resistance to the selective antibiotic associated to any of the tested corresponding expression vectors, but did not express FOXA1. This indicated that the expression of FOXA1 was extremely deleterious for MDA::ER cells. We therefore shifted to an inducible expression system (Tet^{Off}), and generated MDA::ER::FOXA1 and associated control cells (MDA::ER::pCDNA). As evaluated by Western blot and RT-qPCR experiments (**Figure 1A and 1B**), we selected a MDA::ER::FOXA1 clone expressing similar amounts of FOXA1 than MCF-7 cells, which are largely lower than what was observed in the previously characterized MDA::FOXA1 (Serandour et al., 2011) or other MDA::ER::FOXA1 clones.

Interestingly, following the induction of FOXA1 expression during 36h and a 4h treatment with the hormone, the observed E2-sensitive transcriptomes of both control MDA::ER::pCDNA and MDA::ER::FOXA1 cells had little overlap with that of MCF-7 cells (or T47D; data not shown) (**Figure 1C**; see also Supplemental **Figure S1**). There was even a little more common genes between MCF-7 and the control cells than with the MDA::ER::FOXA1 clone (104 vs. 87). The expression of FOXA1 in these cells therefore seemed to engage a distinct transcriptional E2-sensitive program comprising fewer genes than in the other tested cells. This is reinforced by the observation that the presence of FOXA1 provoked the response of 193 novel genes to E2 and abrogated that of 369 genes. RT-qPCR using separately prepared samples validated the pertinence of the microarray-based transcriptome analysis (**Figure 1D**).

Classifying the genes expression levels upon their changes following FOXA1 expression and E2 treatment (**Figure 2A and B**) further indicate that FOXA1 reprogram the E2-sensitive transcriptomes through processes affecting either basal or E2-regulated levels of expression of these genes. Note that only 4 mRNAs exhibited an inversed E2-regulation following the expression of FOXA1 (Inv). Functional annotations of genes that either lose or gained a response to E2 (L.I to L.IV or N.I to N.IV in **Figure 2B**, respectively) clearly showed that FOXA1 promoted the estrogenic regulation of pathways involved in cell apoptosis, and inhibited membrane-associated responses to extracellular signals (**Figure 2C**). Such a reorientation of cell fate was not so obvious for genes whose expression was regulated by FOXA1, either in the presence of E2 or not (**Figure S2**). This indicated that these novel E2-regulated pathways resulted from combined actions of FOXA1 and ER. Furthermore, an E2-dependent drop of the cell population in S phase together with a significant increase of the percentage of cells in G0/G1 was observed by FACS (**Figure 2D**), associated with an increase in the subG1 sub-population of MDA::ER::FOXA1 cells (**Figure 2E**). These functional data altogether therefore indicate that expressing FOXA1 in MDA::ER cells has a net negative effect on cell growth.

Expression of FOXA1 in MDA::ER cells disrupts ER cistrome. We next sought to identify whether the influence of FOXA1 on the E2-sensitive transcriptome of MDA::ER cells could be correlated with an impact on ER

mobilization. To do so, we performed anti-ER ChIP-seq experiments using chromatin prepared from MCF7, MDA::ER::pCDNA and MDA::ER::FOXA1 cells treated for 50 min with E2. As shown in **Figure 3A**, 4,474 ER sites were overlapping in the three cell lines. In accordance with the little correlation observed between MCF-7 and MDA::ER::FOXA1 E2-responsive transcriptomes, only 309 from the 1,655 FOXA1-dependent ER sites (not detected in the MDA::ER::pCDNA control) were common with MCF-7 ERBSs. Qualitatively, the distributions of the identified ERBSs were similar in all three cell lines, with a predominant localization at distant regulatory elements (**Figure S3**). Quantitatively, however, the levels of enrichment of ERBSs in ER were significantly different in MDA::ER::FOXA1 as compared to MDA::ER::pCDNA control cells, as evidenced in the heatmaps of **Figure 3B** and the distribution and mean signals plotted in **Figure 3C**. Unexpectedly, these data therefore indicate that the expression of FOXA1 disrupts the binding of ER onto its cognate sequences when ectopically expressed. This opposes to the described physiological pioneering role of FOXA1. We assessed whether this disruption could be related with a direct binding of FOXA1 on these sites, and whether the few novel ER BSs actually recruited FOXA1 protein. Aligning the signals obtained in anti-FOXA1 ChIP-seq experiments (cf. **Figure S4** for functional annotations) performed in MDA::ER::FOXA1 cells on cell-specific or common ERBSs indicated that FOXA1 is recruited to only 4% of lost ERBSs (**Figure 3D**). In contrast, it was detected on 28 % of ERBSs conserved between control and FOXA1 expressing MDA::ER cells and 41.4 % of the novel ERBSs (see example sites within **Figure 3E**). This percentage was similar to the overlap detected in MCF-7 between ER and FOXA1 ERBSs (**Figure 3D**). Furthermore, as in MCF-7 cells, and published elsewhere in these cells (Hurtado et al., 2011) the binding intensity for ER was almost similar for ER/FOXA1 shared or ER unique regions in MDA::ER::FOXA1 cells (**Figure 3F**). Altogether, these data indicate that the expression of FOXA1 in MDA::ER cells reduces the binding frequency of ER on its sites towards levels that tend to be globally similar to the low enrichment in ER detected on the new ER binding sites. This effect of FOXA1 appears to be indirect, either by acting on the expression levels of another TF or by interfering with the three-dimensional network of interactions that link ERBSs (Fullwood et al., 2009).

Ectopic expression of FOXA1 has a dual impact on H3 K4me2 and K27ac HPTMs. The results obtained so far indicated that the ectopic expression of FOXA1 in MDA::ER cells did not drastically change the cartography of ER binding at genome-scale. We thus wondered whether the FOXA1 BSs (either alone or FOXA1/ER sites) were actually functional and performed H3K4me2 and H3K27ac ChIP-seq experiments to characterize their relative enrichment in histone marks reflective of active or prepared enhancers (cf. **Figure S5** and **S6** for genomic and functional annotations). In both MDA::ER::FOXA1 and MCF-7 cells, the FOXA1/ER sites were apparently more enriched in H3K27ac and H3K4me2 marks than FOXA1 or ER only BSs (**Figure 4A**). Concordant with observations made by others (He et al., 2010; He et al., 2012), the patterns of H3K4me2 signals in MCF-7 cells were either centered over the ERBSs FOXA1 positive or not, or exhibiting two sharp peaks flanking FOXA1 only regions. A similar observation was made for the FOXA1 only sites in MDA::ER::FOXA1 cells. Profiling average H3K27ac signals on the same sites (**Figure 4A**) allowed us to refine these observations that reflect nucleosome dynamics around FOXA1 positive or negative ERBSs. Indeed, in all cases, these signals were clearly bimodal around the regions bound by both or either FOXA1 and ER. Note that the global profiles obtained on all ERBSs were similar to this bimodal distribution of H3K27 acetylation vs. monomodal enrichment of H3K4me2 (**Figure**

S7). As summarized by the sketches and MDA::ER::FOXA1 examples shown in **Figure 4B**, in both FOXA1 positive cell lines, the putative central nucleosome of FOXA1 BSs appears to be less dimethylated at the H3K4 site when it does not concomitantly recruits ER. Since FOXA1 BSs exhibited the same characteristics in MCF-7 and MDA::ER::FOXA1 cells, we concluded that the FOXA1 BSs were likely to be functional in these latest cells.

However, as illustrated within **Figure 4C**, the enrichment of ER or FOXA1 unique and combined sites in H3K4me2 was systematically lower in MDA::ER::FOXA1 cells as compared to control MDA::ER::pCDNA or naïve MDA-MB231 cells. In contrast, the expression of FOXA1 did not significantly alter the content of the three sites categories in the H3K27ac mark whilst ER expression appeared to enhance the acetylation of this residue on histones surrounding its binding sites as compared to MDA-MB231 cells (**Figure 4C**, see also **Figure S7**). Note that such differences were also observed, even with greater amplitudes, when using signal files which were not processed through an inter-sequencing normalization pipeline (**Figure S8**). In order to understand more precisely the actions of FOXA1 on the levels of H3K4 dimethylation and K27 acetylation on shared ER/FOXA1 sites, we aligned the HPTMs ChIP-seq signals on conserved, novel or lost ERBS as defined previously (**Figure 3**). As shown within the **Figure 4D**, novel ERBS that are shared FOXA1 sites in MDA::ER::FOXA1 cells appeared to be slightly more enriched in H3K4me2 and clearly exhibited a bimodal H3K27ac signal which was not observed in either MDA-MB231 or control MDA::ER::pCDNA cells. This is concordant with the hypothesis that these ERBS acquire their functionality in a FOXA1 manner. In contrast, lost ERBSs or shared ER/FOXA1 sites contained less H3K4me2 following the expression of FOXA1 whilst their mean enrichment in H3K27ac appeared unchanged. However, lost ERBSs exhibited a tendency to lose their bimodal distribution in H3K27ac (**Figure 4D**). These data therefore show that FOXA1 expression in MDA::ER cells has a global negative impact on the H3K4me2 content of genomic regions that it shares with ER (**Figure 4D**) or not (FOXA1 only sites in **Figure 4C**). But in parallel, although on a limited number of regions, it promotes the functionality of novel ERBSs specific for MDA::ER::FOXA1 cells. Finally, we observed that ER and FOXA1 BSs were presenting significant levels of H3K4me2 or of monomodal H3K27ac signals in MDA-MB231 cells (**Figures 4C, 4D, S7 and S8**). This implies that ectopically expressed ER and FOXA1 have a certain tendency to bind to regions that already contain marks for active/poised chromatin.

FOXA2 is recruited on a subset of MDA::ER ERBS. Our transcriptome analyses have shown that the expression of FOXA1 in MDA::ER cells reinforces the inhibitory effect of E2 on cell growth, contrasting with the proliferative impact of this hormone in MCF-7 cells. This is in accordance with a recent study that has also demonstrated that another TF, GATA3, might be required in addition to FOXA1 to reconstitute an E2-responsive phenotype in ER negative cell lines (Kong *et al.*, 2011). However, contrasting with the GATA3 negative cell line used in that study, our MDA::ER cells do express this TF, although to lower levels than MCF-7 cells (Quintin *et al.*, submitted). Moreover, we found GATA3 to collaborate with other GATAs for directing ER actions in this particular cell line (Quintin *et al.*, submitted). In addition, the fact that FOXA1 expression reduced the binding of ER (**Figure 3**) and a concomitant negative impact on the enrichment of ER and FOXA1 BSs in H3K4me2 HPTM led us to hypothesize that it may compete or interfere with the actions of an undetermined TF on these BSs. Interestingly, besides identifying motifs for “classical” collaborative factors within

MDA::ER::pCDNA ERBS, bioinformatics analyses using two different algorithms (SeqPos, He *et al.*, 2010 ; and CentDist, Zhang *et al.*, 2011) also identified Forkhead motifs as being enriched in these cells that do not express FOXA1 (**Figure 5A**). Moreover, we found that FOXA2 exhibited an expression pattern inversely correlated with that of FOXA1: highly expressed in MDA::ER cells but not in MCF7 cells, and *vice-versa* (**Figure 5B and Figure S9**). As determined following ChIP-seq experiments, approximately 30 % of MDA::ER::pCDNA ERBSs were also detected as FOXA2 positive (**Figure 6C**). As previously observed with FOXA1 in MCF-7 and MDA::ER::FOXA1 cells, the distribution of the H3K4me2 signal on FOXA2only BSs was bimodal in MDA::ER::pCDNA cells; as was the case for shared ER/FOXA2 BSs although to a lesser extent (**Figure 6D**). The mean enrichment of ER/FOXA2 BSs in H3K27ac was also found to be higher than for ER or FOXA2 unique sites (**Figure 6D**). Combined with the other observations, these ChIP-seq data indicate that FOXA2 might be an important collaborating factor for ER in MDA::ER::pCDNA cells.

FOXA2 behaves as a ER pioneering factor in MDA::ER cells. We next aimed at determining whether FOXA2 could play in MDA::ER a similar role than the one exerted by FOXA1 in MCF-7 cells. We transfected MDA::ER::pCDNA cells with siRNAs directed against FOXA2 expression for 36h and mapped ER binding sites by ChIP-seq following E2 treatment. We used a siRNA directed against the *Luciferase* (siLuc) mRNA as a control for these experiments. As depicted within **Figure 6A**, there was no detectable change in ER mRNA levels following the transfection of FOXA2 siRNAs, but ER binding to its targets was significantly altered. This was true in both quantitative (loss of almost 50% of ERBSs **Figure 6B**) and qualitative manner (decrease in ER signal intensities of more than 90% of ERBSs in **Figure 6C and 6D**), as exemplified in **Figure 6E**. We also identified 435 novel ERBSs in siFOXA2 transfected cells (**Figure 6B**). As shown for FOXA1 in MCF-7 cells (Hurtado *et al.*, 2011), contrasting with the drastic effects observed on ER binding following the reduction in FOXA2 amounts, only 30 % of ERBSs of MDA::ER cells actually mobilized FOXA2 (**Figure 5C**). This overlap was almost similar for lost ERBSs (**Figure 6F**); but only 16.1% of ER binding sites that were conserved between control and siLuc or 5.5% of the novel ones recruited FOXA2 (**Figure 6F**). These data suggest that FOXA2 is indirectly regulating ER binding on a number of its BSs, either through long-distance interactions between BSs or through regulating the expression or function of a secondary factor that consequently regulates ER binding.

To further characterize the requirement of FOXA2 for the correct establishment of an E2-sensitive regulatory network, we evaluated the impact of reducing the amounts of this protein on the ER-mediated regulations of the TFF cluster (**Figure 7**. From the 17 ERBS that we previously identified by ChIP-chip on this genomic region (see Quintin *et al.*, submitted; all sites confirmed here in our ChIP-seq experiments, data not shown), 9 were found to also recruit FOXA2 (**Figure 7A**). Silencing FOXA2 in MDA::ER::pCDNA cells reduced ER binding to 8 of these sites, as well as the relative enrichment in H3K4me2 and H3K27ac HPTMs of 7 or 8 of these ERBSs, as evaluated by ChIP-qPCR assays (**Figure 7B**). 3C-qPCR experiments next demonstrated that FOXA2 was necessary for achieving a maximum of interactions linking these ER/FOXA2 BSs to the promoters of estrogen-sensitive genes (**Figure 7C**). The final output of these disruptions in FOXA2 and ER recruitments was the significant reduction of the fold inductions of the E2-sensitive genes of this genomic loci, except *RIPK4* and *TMPRSS3* (**Figure 7D**). We can conclude from all these analyses that FOXA2 is required for a proper mediation

of E2 signalling to its target genes through ER, and that this protein appear to act as a pioneering factor for ER in the MDA::ER cell system.

A FOXA1/FOXA2 competition reduces ERBSs' functionality. This functional overlap between FOXA2 and ER BSs and the fact that 60% of the FOXA1 BSs detected in MDA::ER::FOXA1 cells were bound by FOXA2 in control MDA::ER::pCDNA cells (not shown) lead us to question whether FOXA2 was interfering with the action of ectopically expressed FOXA1 in MDA::ER::FOXA1 cells. We first verified by ChIP experiments that regions bound by FOXA1 in MCF-7 cells were not enriched by the anti-FOXA2 antibody, which was also false for the converse experiment in MDA::ER::pCDNA cells (data not shown). The overlap between the cistromes of the two FOXA proteins is therefore unlikely to be generated by artefactual cross-reactions of antibodies. To explore the competition hypothesis, we mapped FOXA2 cistrome in MDA::ER::FOXA1 cells following the induction of FOXA1 expression and found that approximately 55 % of the 23,903 FOXA2 BSs determined in these cells were actually also FOXA1 positive (**Figure 8A**). Re-ChIP experiments (not shown) confirmed on randomly chosen sequences these overlaps. We next classified MDA::ER::FOXA1 ERBSs depending upon their apparent co-recruitment of FOXA1 and or FOXA2 and plotted ER enrichments on these sites measured in MDA::ER::FOXA1 against those measured in control MDA::ER::pCDNA cells (**Figure 8B**). This lead us to categorize genomic regions that exhibited enhanced (green dots), unchanged (grey) or reduced (red) enrichments in ER. The analyses of ChIP-seq signals obtained for FOXA proteins and modified histones on these sites (**Figure 8C**) made apparent that ER/FOXA1 sites on which ER binding was enhanced actually exchanged a FOXA2 binding observed in control cells for a mobilization of FOXA1. This exchange was accompanied with an apparent tendency of these sites to gain H3K4me2 modification. In contrast, there was no FOXA2/FOXA1 exchange for the two other categories of ER/FOXA1 sites. This indicates that FOXA1 is likely to have a stronger influence on chromatin than FOXA2 on these sites. These analysis also showed that in the case of combined ER/FOXA/FOXA2 sites, those that exhibited a significant (>2 fold) reduction in ER binding (red dots in **Figure 8B**) were also mobilizing less FOXA2 (**Figure 8C**, extreme right panels). This suggests that FOXA1 binding on these sites may compete for FOXA2 actions and interfere with the pioneering or collaborating actions of this latter protein on ER mobilization, as hypothesized. On the other hand, there was no striking variations in H3K4me2 or H3K27ac levels on these sites. To finally validate the link between the FOXA1/FOXA2 competition and ER-mediated signalling, we aimed at silencing FOXA2 expression before the induction of FOXA1 in MDA::ER::FOXA1 cells (**Figure 8D**). We then analysed events occurring onto to ER/FOXA1/FOXA2 BSs on which the FOXA1/FOXA2 competition reduced the binding of ER. As shown in **Figure 8E**, silencing FOXA2 expression restored or enhanced ER binding to these tested sites, associated -except for the second one- with a greater association of FOXA1 and histone modifications. These latest results validate the pertinence of our initial hypothesis linking a competition between FOXA1 and FOXA2 to altered recruitments of ER. It remains to be tested whether this mechanism is responsible for the absence of re-orientation of MDA::ER cells towards cell proliferation upon FOXA1 ectopic expression.

DISCUSSION

FOXA1 has been determined to be an essential component of ER signalling, and eliciting breast cancer cells proliferation (Perou et al., 2000). In contrast, we show here that the ectopic expression of FOXA1 in the MDA::ER system has a negative influence on cell proliferation, as also observed by others (Kong et al., 2011). But in addition, we found that FOXA1 reduced both the number of ER binding events and their frequency in these cells. At first sight, these observations were apparently opposed to the well-described pioneering action of FOXA1 on directing ER cistrome (Carroll et al., 2006; Hurtado et al., 2011). When going into a further detailed analysis of our ChIP-seq and transcriptome data, we however found approximately 650 novel ERBSs in the MDA::ER::FOXA1 cells which were also FOXA1 positive. Importantly, the distribution of the histone marks present at these sites were concordant with functional enhancers with bimodal H3K27ac signals (He et al., 2010; He et al., 2012). We also observed that this functionalization was also associated on some sites (4 tested so far, data not shown) with DNA undergoing coincident demethylation and hydroxymethylation processes when comparing FOXA1 negative MDA and MDA::ER cells to MDA::ER::FOXA1, as it might be expected from our previous work (Serandour et al., 2011; Serandour et al., 2012). However, these novel ERBSs were not found at the proximity of genes acquiring or losing their E2 responsiveness (data not shown). These specific changes in the transcriptome of MDA::ER cells provoked by the expression of FOXA1 are therefore likely to be associated with the negative influence of this protein onto ER-mediated regulations.

A recent work has given support for the GATA3 factor to be an essential component for the formation of ER/FOXA1/GATA3 enhanceosomes to restore MDA-MB231 cell growth and ER-mediated transcription in these cells (Kong et al., 2011). As we found that our MDA::ER cells were expressing GATA3, although to lower levels than MCF-7 cells (Quintin et al., submitted), we deduced that there was an additional player in our cells that might hinder FOXA1 actions. Searches for enriched motifs for TF binding in the MDA::ER ERBSs DNA surprisingly identified Forkhead motifs, leading us to characterize FOXA2 as a major candidate for this hindering protein. Our whole-genome analysis of FOXA2 binding to chromatin led us to hypothesize that this forkhead protein might constitute an alternate pioneer factor for ER in the MDA::ER cell system. This hypothesis was substantiated by demonstrating that FOXA2 was required for the ER-mediated estrogenic-regulations of the TFF cluster of genes. FOXA2 has a high homology with FOXA1 (Lai et al., 1991) and has the ability to regulate transcription (Gao et al., 2008; Wederell et al., 2008). We showed, at the TFF locus scale, that FOXA2 was required for the recruitment of ER and associated chromatin modifications (H3K4me2 and H3K27ac). Whether this is true at the genome scale, indicating that this protein may exhibit also the same properties of FOXA1 towards chromatin remodeling (Cirillo et al., 2002) remains however to be determined. Of note, it has been demonstrated in liver that the binding of FOXA2 was regulated by the H3K4me1 status of the nucleosomes present around its BSs (Hoffman et al., 2010), reminiscent with the requirement of H3K4 modifications for FOXA1 binding (Lupien et al., 2008). We next reasoned that if FOXA2 was a pioneering factor in MDA::ER cells, the supplemental expression of FOXA1 might interfere with these actions. Or, conversely, that FOXA2 may interfere with FOXA1-mediated re-orientation of cell phenotypes towards proliferation. Such complex interrelations between FOXA proteins have been observed in a more "natural" cell context (hepatocellular

carcinoma HepG2 cells), with FOXA1, FOXA2 and FOXA3 able to i) interact with each other except FOXA1 and FOXA3; ii) bind to similar genomic regions; iii) have specific cisomes (Motallebipour et al., 2009). Furthermore, an opposite role of FOXA1 and FOXA2 was recently determined in liver for the regulation of the *DIO1* gene (Kanamoto et al., 2012). Interestingly, although yet tested only on a few sites, we observed that silencing FOXA2 before inducing the expression of FOXA1 allowed a stronger binding of this latter protein, associated with an increased binding of ER. This is in line with our hypothesis of an interfering competition between FOXA1 and FOXA2 for ER binding, with FOXA1 likely having a more potent pioneering action than FOXA2 when it is able to bind to chromatin. These observations have to be generalized by considering more sites, such as the FOXA1/FOXA2 category. We now have to test whether this interfering interrelation between endogenous FOXA2 and ectopically expressed FOXA1 is actually blocking a reprogramming of MDA::ER cells towards a proliferative phenotype. This will be stressed through RT-qPCR experiments testing the expression of genes involved in cell proliferation following the transfection of siFOXA2 before inducing FOXA1 in MDA::ER::FOXA1 cells, and through measurements of cell proliferation (cell counting and FACS analysis). Furthermore, although expressed in our cells, we also have to consider the effect of GATA3 expression in our cells. We previously postulated that GATA3 was controlling FOXA1 actions on ERBSs, in accordance with what was found on a subset of ERBS in a recent publication (Theodorou et al., 2013). The questions that remain to be considered are now to know whether i) GATA3 overexpression displaces FOXA2 to allow a FOXA1-mediated reorientation of MDA::ER phenotype observed by others (Kong et al., 2011); and ii) GATA3 also control FOXA2 binding and subsequent chromatin remodelling. Importantly, we recently performed pilot RT-qPCR experiments following the reduction of GATA3 in our MDA::ER::FOXA1 cells and did not observe any effect on ER or FOXA2 mRNA levels. This suggests that these three proteins are not *stricto sensu* constituting a functional network in these cells.

Importantly, we demonstrated in this work that the overexpression of FOXA1 in MDA::ER cells provoked a strong reduction in H3K4me2 levels associated with a change in the distribution of the H3K27ac signal around ER BSs and a subset of FOXA2 sites. We also performed pilot experiments showing that these negative effects on the chromatin status of ERBSs were for some sites going beyond histone modifications, with significant increase in DNA methylation and decrease in DNA hydroxymethylation (data not shown). Besides these observations, we also showed that the majority of FOXA1 negative impact on ER binding was likely indirect: only 4% of the lost ERBSs actually gained the ability to mobilize FOXA1. This could reflect the fact that many ERBSs are linked through chromatin loops (Fullwood et al., 2009). Hence, events occurring on one FOXA1/ER site might have consequences on other ERBSs. Alternatively, or additionally, this also suggests that i) the FOXA1/FOXA2 interrelation might target a common histone methylase to chromatin; ii) that at least FOXA1 may control the expression of this histone modifier, explaining its indirect effects; and iii) DNA methylases and de-hydroxymethylation events may also be involved. Interestingly, such indirect effects are also suggested by the observed reductions in ER binding in MCF-7 cells following the reduction in FOXA1 amounts by siRNAs on sites that did not recruit FOXA1 (Hurtado et al., 2011). We do have candidate proteins to consider: we found that the expression of the histone methyltransferases MLL2, MLL3, MLL5, SMYD2 and SMYD3 were affected in

MDA::ER::FOXA1 as compared to control MDA::ER::pCDNA cells. Interestingly, SMYD3 was characterized as a ER coactivator in MCF-7 cells (Kim et al., 2009) as was a JMJD2B/MLL2 complex (Shi et al., 2011).

In summary, our results indicate that the expression of FOXA1 in a cellular context in which FOXA2 is acting as a pioneering factor for ER mediated regulation of transcription is interfering with ER binding through mechanisms that apparently involve chromatin modifiers. The functional output of these complex interrelations is an orientation of cells towards apoptosis or at least a reduction of cell proliferation that opposes to classically defined estrogenic response of ER positive breast cancer cells. It remains to be tested whether a reduction in FOXA2 amounts is sufficient to allow a concerted reorientation of MDA-MB231 cells towards an estrogen-sensitive growth phenotype by FOXA1 and ER, or not.

ACCESSION NUMBERS

The data generated during this study are going to be deposited in NCBI's Gene Expression Omnibus (Barrett et al., 2009).

ACKNOWLEDGEMENTS

This work was supported by the CNRS, the University of Rennes I, and benefited from grants of the ARC, the Ligue Contre le Cancer (Equipe Labellisée Ligue 2009), Région Bretagne (CREATE 4793) and the the Agence Nationale pour la Recherche (ANR-09-BLAN-0268-01). J.Q. is a recipient of a joint fellowship by the CNRS and the Region Bretagne.

AUTHOR CONTRIBUTIONS

All biological experiments performed by JQ, CLP, MB, GP, FP and RM were conceived and designed by RM, JQ and GS. SA programmed all local informatic tools to analyze whole-genome data. RM and JQ wrote the manuscript with the help of all other authors.

REFERENCES

- Aridsson S, Kwasniewski M, Riano-Pachon DM, Mueller-Roeber B. (2008). QuantPrime--a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics*. 9, 465.
- Barrett T, Troup DB, Wilhite SE, Ledoux P., Rudnev D., Evangelista C., Kim I. F., Soboleva A., Tomashevsky M., Marshall K. A., et al. (2009). NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res*. 37, D885-890.
- Blankenberg D, Taylor J, Schenck I, He J, Zhang Y, Ghent M, Veeraraghavan N, Albert I, Miller W, Makova KD, et al. (2007). A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. *Genome Res* 17, 960-964.
- Carroll JS, Meyer CA., Song J, Li W, Geistlinger TR, Eeckhoutte J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, et al. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nat Genet*. 38, 1289-1297.
- Carroll JS, Brown M. (2006). Estrogen receptor target gene: an evolving concept. *Mol. Endocrinol*. 20, 1707-1714.
- Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, Zaret KS. (1998). Binding of the winged helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J*. 17, 244-254.
- Cirillo L.A et Zaret K.S (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Molecular Cell*. 4, 961-969.
- Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 9, 279-289.
- Duncan DT, Prodduturi N, Zhang B. (2010). WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics*. 11:P10.
- Eeckhoutte J, Métiévier R, Salbert G. (2009). Defining specificity of transcription factor regulatory activities. *J Cell Sci*. 122, 4027-4034.
- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 462, 58-64.
- Gao N, LeLay J, Vatamaniuk MZ, Rieck S, Friedman JR, Kaestner KH. (2008). Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev*. 22, 3435-3448.
- Garcia M, Derocq D, Freiss G, Rochefort H. (1992). Activation of estrogen receptor transfected into a receptor-negative breast cancer cell line decreases the metastatic and invasive potential of the cells. *Proc Natl Acad Sci U S A*. 89, 11538-1142.
- Gondor A, Rougier C, Ohlsson R. (2008). High-resolution circular chromosome conformation capture assay. *Nat Protoc* 3, 303-313.
- Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2, 1722-1733.
- Hall JM, Couse JF, Korach KS. (2001) The multifaceted mechanisms of estradiol and estrogen receptor signaling. *J Biol Chem*. 276, 36869-36872.
- Hoffman BG, Robertson G, Zavaglia B, Beach M, Cullum R, Lee S, Soukhatcheva G, Li L, Wederell ED, Thiessen N,

- et al. (2010). Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. *Genome Res.* 20, 1037-1051.
- Kanavos P. (2006). The rising burden of cancer in the developing world. *Ann Oncol.* 17 Suppl 8:viii15-viii23.
- Kanamoto N, Tagami T, Ueda-Sakane Y, Sone M, Miura M, Yasoda A, Tamura N, Arai H, Nakao K. (2012). Forkhead box A1 (FOXA1) and A2 (FOXA2) oppositely regulate human type 1 iodothyronine deiodinase gene in liver. *Endocrinology.* 153, 492-500.
- Kim H, Heo K, Kim JH, Kim K, Choi J, An W. (2009). Requirement of histone methyltransferase SMYD3 for estrogen receptor-mediated transcription. *J Biol Chem.* 284, 19867-19877.
- Kong SL, Li G, Loh SL, Sung W-K, Liu ET. (2011). Cellular reprogramming by the conjoint action of ER α , FOXA1, and GATA3 to a ligand-inducible grpwth state. *Mol Syst Biol.* 7, 526.
- Langmead B, Trapnell C, Pop M, Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. (2009) The Sequence alignment/map format and SAMtools. *Bioinformatics.* 25, 2078-2079.
- Lai E, Prezioso VR, Tao WF, Chen WS, Darnell JE Jr. (1991). Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head. *Genes Dev.* 5, 416-427.
- Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell.* 132, 958-970.
- Magnani L, Eeckhoute J, Lupien M. (2011a). Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet.* 27, 465-474.
- Magnani L, Ballantyne EB, Zhang X, Lupien M. (2011b). PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS Genet.* 7:e1002368.
- McEwan IJ. (2009). Nuclear receptors: one big family. *Methods Mol Biol.* 505, 3-18.
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495-501.
- Metivier R, Penot G, Hubner MR, Reid G, Brand H, Kos M, Gannon F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell.* 115, 751-763.
- Mohn F, Weber M, Schubeler D, Roloff TC. (2009). Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol.* 507, 55-64.
- Motallebipour M, Ameer A, Bysani MSR, Patra K, Wallerman O, Mangion J, Barker MA, McKernan KJ, Komorowski J, Wadelius C. (2009). Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq. *Genome Biol.* 10: R129.
- Nott SL, Huang Y, Li X, Fluharty BR, Qiu X, Welshons WV, Yeh S, Muyan, M. (2009). Genomic responses from the estrogen-responsive element-dependent signaling pathway mediated by estrogen receptor alpha are required to elicit cellular alterations. *J Biol Chem.* 284, 15277-15288.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA et al. (2000). Molecular portraits of human breast tumours. *Nature.* 406, 747-752.

- Quintin J, Le Péron C, Palierne G, Bizot M, Cunha S, Sérandour AA, Avner S, Percevault F, Huet S, Watrin E, Eeckhoutte J, Legagneux V, Salbert G, Métivier R. Cell-specific and spatio-temporal controls of the estrogen-responsive trefoil factor (TFF) locus activity. *Mol. Cell*, Submitted.
- Reid G, Hubner MR, Metivier R, Brand H, Denger S, Manu D, Beaudouin J, Ellenberg J, Gannon F. (2003). Cyclic, proteasome-mediated turnover of unliganded and liganded ERalpha on responsive promoters is an integral feature of estrogen signaling. *Mol Cell*. 11, 695-707.
- Russo J, Russo H. (2006). The role of estrogen receptor in the initiation of breast cancer. *J. Steroid Biochem. Mol. Biol*. 102, 89-96.
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M. et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374-378.
- Serandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Metivier R, et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res*. 21, 555-565.
- Serandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierne G, Gheeraert C, Barloy-Hubler F, Péron CL, et al. (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res*. 40, 8255-8265.
- Shi L, Sun L, Li Q, Liang J, Yu W, Yi X, Yang X, Li Y, Han X, Zhang Y, Xuan C, Yao Z, Shang Y. (2011). Histone demethylase JMJD2B coordinates H3K4/H3K9 methylation and promotes hormonally responsive breast carcinogenesis. *Proc Natl Acad Sci U S A*. 108, 7541-7546.
- Tan SK, Lin ZH, Chang CW, Varang V, Chang KR, Pan YF, Yong EL, Sung WK, Cheung E. (2011). AP-2γ regulates oestrogen receptor-mediated long-range chromatin interaction and gene transcription. *EMBO J*. 30, 2569-2581.
- V, Stark R, Menon S, Carroll JS. (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res*. 23, 12-22.
- Wederell ED, Bilenky M, Cullum R, Thiessen N, Dagpinar M, Delaney A, Varhol R, Zhao Y, Zeng T, Bernier B, et al. (2008). Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res*. 36, 4549-4564.
- Welboren WJ, Sweep FC, Span PN, Stunnenberg HG. (2009). Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated? *Endocr. Relat. Cancer*. 16, 1073-1089.
- Zaret KS, Carroll JS. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 25, 2227-2241.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson WE, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 9, R137.

FIGURE LEGENDS

Figure 1. FOXA1 expression reprograms the estrogen-sensitive transcriptome of MDA::ER cells and strongly reduces their proliferation rate. The expression of FOXA1 (inducible) and ER (constitutive) proteins was verified by Western blot **(A)** and RT-qPCR experiments **(B)** in our different cellular clones derived from naive ER and FOXA1 negative MDA-MB231 cell line. FOXA1 expression was induced by a 48h treatment with tetracyclin (Tet) at a concentration of 1.5 µg/ml in culture media. The expression values detected for *FOXA1* and *ER* mRNAs were reported to those of the invariant *Rplp0* control gene, whilst equivalent loading of the blotted gels was controlled by β-actin amounts. WM lane stands for weight markers. The FOXA1 and ER protein and mRNAs levels measured in MCF-7 cells served as reference to select a MDA::ER::FOXA1 clone expressing similar amounts of both TFs (boxed lanes in panel A). **(C)** Venn diagram illustrating the overlap between the E2-sensitive transcriptomes of the control MDA::ER::pCDNA and MDA::ER::FOXA1 cell lines. These were determined following a 48h treatment with Tet combined with a 4h treatment with 10^{-8} M E2 or ethanol (EtOH, vehicle control), and are compared with that of MCF-7 cells (previous GSE23850 data obtained in the laboratory). **(D)** Heatmaps summarizing results from RTqPCR experiments controlling the cell-specific fold induction -as defined in the microarrays data- of indicated genes by E2 in the three cell lines. N.D stands for not determined. Shown are mean log2 fold changes calculated from values obtained in three independent triplicate experiments.

Figure 2. FOXA1 reduces the growth rate of MDA::ER cells. **(A)** Venn diagram comparing the MDA::ER::FOXA1 genes populations exhibiting significant changes in their expression levels as compared to MDA::ER::pCDNA cells in the absence or presence of E2. **(B)** Heatmaps of a K-means clustering of the expression levels of mRNAs regulated by E2 either only in control MDA::ER::pCDNA cells (lost in MDA::ER::FOXA1 cells: L), only in MDA::ER::FOXA1 cells (Novel: N), or in both cells (Common: C). Profiles for common genes exhibiting an inverse regulation (Inv) are depicted at the bottom of the panel. Lines on the right schematize the observed variations in expression for each of the 4 clusters defined for the three L, N and C categories. **(C)** Top ten functional annotations of the genes belonging to the L, N and C groups. The Top ten pathways (compilations of Kegg, common and wiki pathways) are also indicated with associated Benjamini&Hochberg (BH) corrected *p*-value. **(D and E)** FACS analysis of indicated cell lines following the induction of FOXA1 or not (plus or minus Tet) and in the presence or not of E2 10^{-8} M for 4h.

Figure 3. Cell-specific cistromes of ER and FOXA1. **(A)** Venn diagrams of ER binding sites (BS) detected by ChIP-seq in MCF-7, MDA::ER::pCDNA and MDA::ER::FOXA1 cells treated by E2 for 50 min. **(B)** Heatmap showing the k-means ranked signal intensities of ER binding in all three cell lines in a -/+ 5kb window centered on the ERBSs. **(C)** Distribution of ER binding event intensities, defined as the ER signal detected within the central 100 bp of the ERBSs, on MDA::ER::pCDNA or MDA::ER::FOXA1 specific or common ERBSs. Mean values ± SEM are shown within the insert. Significances of the differing signals were evaluated by unpaired t-test and F-tests (***) and ####: *p*<0.0001 for differences between the enrichment of common ERBSs in the two cell lines or between the signals measured on cell-specific ERBSs in the corresponding cell line). **(D)** Average signal (left side of the panel) or heatmaps of the signal intensities of FOXA1 binding in MDA::ER::FOXA1 cells within -/+ 5kb around ERBSs

classified depending upon their overlap with MDA::ER::pCDNA ones. MCF-7 profiles are given as references at the bottom of the panel. Heatmaps represent FOXA1 binding intensities ranked from the strongest to the lowest signal measured. Numbers of overlapping ER/FOXA1 sites for each categories are indicated on the right side of the heatmap as well as the % of overlap in brackets **(E)** Examples of ER and/or FOXA1 binding regions in the three cell lines. **(F)** Average signal intensity of ER binding to ER/FOXA1 shared regions or to ER only recruiting sequences in a ± 5 kb window .

Figure 4. FOXA1 modulates the chromatin organization at its binding sites. **(A)** Average intensity of H3K4me2 and H3K27ac ChIP-seq signals in MDA::ER::FOXA1 (left) or MCF-7 (right) cells on unique or combined ER and FOXA1 sites. **(B)** Examples of H3K4me2 and H3K27ac enrichments on regions potentially extending around three nucleosomes (Nuc) centered on unique or combined ER and FOXA1 BSs in MDA::ER::FOXA1. In the case of the FOX^{only} BSs, the interrogation mark stands for either a central nucleosome exhibiting no H3K4me2 and H3K27ac marks or its depletion. **(C)** Alignment of H3K4me2 (left) or H3K27ac (right) signals obtained by ChIP-seq in MDA-MB231 (black line), MDA::ER::pCDNA (grey) and MDA::ER::FOXA1 (brown) on unique or combined ER and FOXA1 BSs of MDA::ER::FOXA1 cells. **(D)** Same as in panel C but using ER/FOXA1 BSs categorized for their cell-specificity: new or lost sites in MDA::ER::FOXA1 as compared to control MDA::ER::pCDNA cells (top or bottom of the panel, respectively), or common ER sites that also recruit FOXA1 when it is expressed (middle panel).

Figure 5. FOXA2 recruited to a subset of ERBS in MDA::ER cells. **(A)** Motif searches on ERBSs from MCF-7, MDA::ER::pCDNA and MDA::ER::FOXA1 cells. Screenings were made using two different algorithms (SeqPos and CentDist) for accuracy. Only motifs present within the two databases of results were ranked, and the top10 are shown here. **(B)** RT-qPCR experiments performed in MDA-MB231, MDA::ER and MCF-7 cells assessing for the expression of *FOXA1* and *FOXA2* mRNAs. Histograms shown represent means + SD of values obtained in two different triplicate experiments. Both mRNAs expression were normalized to that of the *Rplp0* control. **(C)** Heatmap representation of the binding signal of FOXA2 and ER in a ± 5 kb window surrounding MDA::ER::pCDNA ERBSs. Binding events are here ranked from strongest to lowest FOXA2 signal. **(D)** Average signals for H3K4me2 (top) or H3K27ac (bottom) on unique or shared ER and FOXA2 BSs in MDA::ER::pCDNA cells.

Figure 6. Recruitment of ER in MDA::ER cells requires FOXA2. **(A)** RT-qPCR experiment controlling the efficiency of the siFOXA2 in depleting its target mRNA but not *ER* one. Controls were made by transfecting a siRNA directed against the *Luciferase* gene (siLuc). Values are mean \pm SD of data normalized to the expression level of the control *Rplp0* mRNA obtained in two independent triplicates (***) : Wilcoxon p-value <0.001). **(B)** Overlap of identified ERBSs in MDA::ER::pCDNA cells transfected with siFOXA2 or control siLuc siRNAs. **(C)** Heatmap showing the intensity of ER binding in siLuc and siFOXA2 transfected MDA::ER::pCDNA cells in a ± 5 kb window around the ERBSs identified in the control siLuc condition. Intensities are ranked from the highest to the lowest signal detected in cells transfected with the control siRNA. **(D)** Scatterplot comparing ER binding signals in siFOXA2 vs. siLuc conditions on ERBSs of the three categories: conserved, lost or novel ones following transfection of the siFOXA2. **(E)** Examples of ER binding in siLuc and siFOXA2 conditions. FOXA2 signal obtained

in untransfected cells is given as a reference. **(F)** Overlap of FOXA2 BSs (left) with conserved, lost and novel ERBSs categories defined earlier. Percentages of overlap are given in brackets. Heatmap on the right show ranked binding intensity of FOXA2 in a -5/+5 kb window centered on these sites.

Figure 7. FOXA2 is required for the estrogenic regulation of the TFF cluster. **(A)** ER and FOXA2 ChIP-seq profiles on the 9 FOXA2 positive ERBS from the TFF cluster (chr.21). Numerotation of the sites respect the nomenclature defined in Quintin et al. (submitted) **(B)** Heatmaps summarizing ChIP-qPCR experiments performed in E2-treated MDA::ER::pCDNA cells transfected for 36h with siRNAs directed against the *luciferase* (Luc, control) or *FOXA2* mRNAs. **(C)** Heatmap of values obtained by 3C-qPCR experiments evaluating the impact of the siFOXA2 on the interaction frequencies between the promoters of the E2-regulated genes included in the TFF cluster and the ER/FOXA2 BSs. The control interactome is depicted at the top of the panel with lines linking ERBSs with associated genes promoters. Lines are drawn either plain or dotted only for a sake of clarity. **(D)** RT-qPCR experiments were performed following the transfection of control siLuc or siFOXA2 in MDA::ER::pCDNA cells and a 4h treatment with E2 or vehicle control. Values were normalized to the expression of the control *Rplp0* gene and fold inductions by E2 were calculated from these data. Shown are mean \pm S.D of 5 values obtained in two independent experiments. Wilcoxon test was used to evaluate the statistical significance of the observed variations (**: $p < 0.01$; ***: $p < 0.001$).

Figure 8. FOXA2 hinders FOXA1 actions . **(A)** Venn diagram depicting the overlap between FOXA2 and FOXA1 binding sites determined in MDA::ER::pCDNA cells and MDA::ER::FOXA1 cells, associated with the heatmap representation of FOXA1 and FOXA2 ChIP-seq signals (right) within a -5/+5 kb window on clustered FOXA1/2 unique or shared binding sites. **(B)** Scatterplot comparing ER binding signals in MDA::ER::FOXA1 vs. MDA::ER::pCDNA cells on regions recruiting either ER and FOXA1 only or all three ER/FOXA1/FOXA2 proteins in MDA::ER::FOXA1 cells. **(C)** Average signals for FOXA proteins, H3K4me2 or H3K27ac on sites of the ER/FOXA1 and ER/FOXA1/FOXA2 categories either gaining (green dots), losing (red) or exhibiting no change (grey) in ER mobilization. **(D)** RNA were prepared from MDA::ER::pCDNA and MDA::ER::FOXA1 cells transfected with siFOXA2 or siLuc as control, and subjected to RT-qPCR assays aiming to measure indicated mRNAs amounts. Results shown are mean \pm S.D of values normalized to those of the *Rplp0* control obtained in three independent triplicate experiments. **(E)** Heatmaps summarizing ChIP-qPCR experiments measuring the relative enrichment of ER/FOXA1/FOXA2BSs in ER, FOXA proteins and H3K4me2 and H3K27ac marks in both control and MDA::ER::FOXA1 cells. Results obtained in two independent triplicate experiments were expressed as relative enrichments towards a nonspecific ChIP (beads alone) and either values obtained on a negative control sequence (ER and FOXAs) or those obtained for a ChIP directed against unmodified H3.

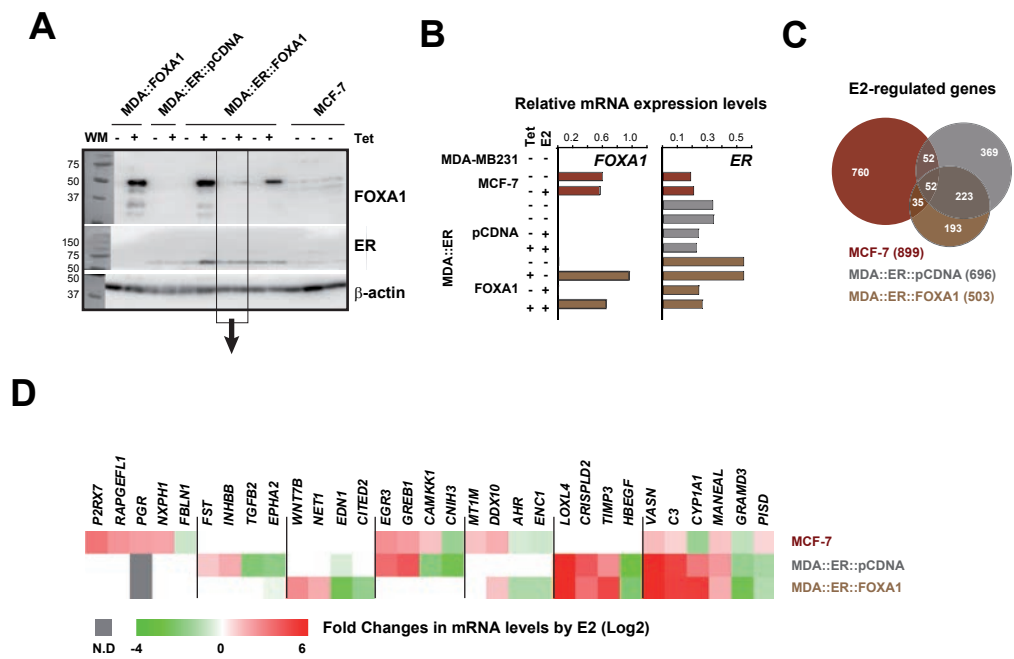


Figure 1

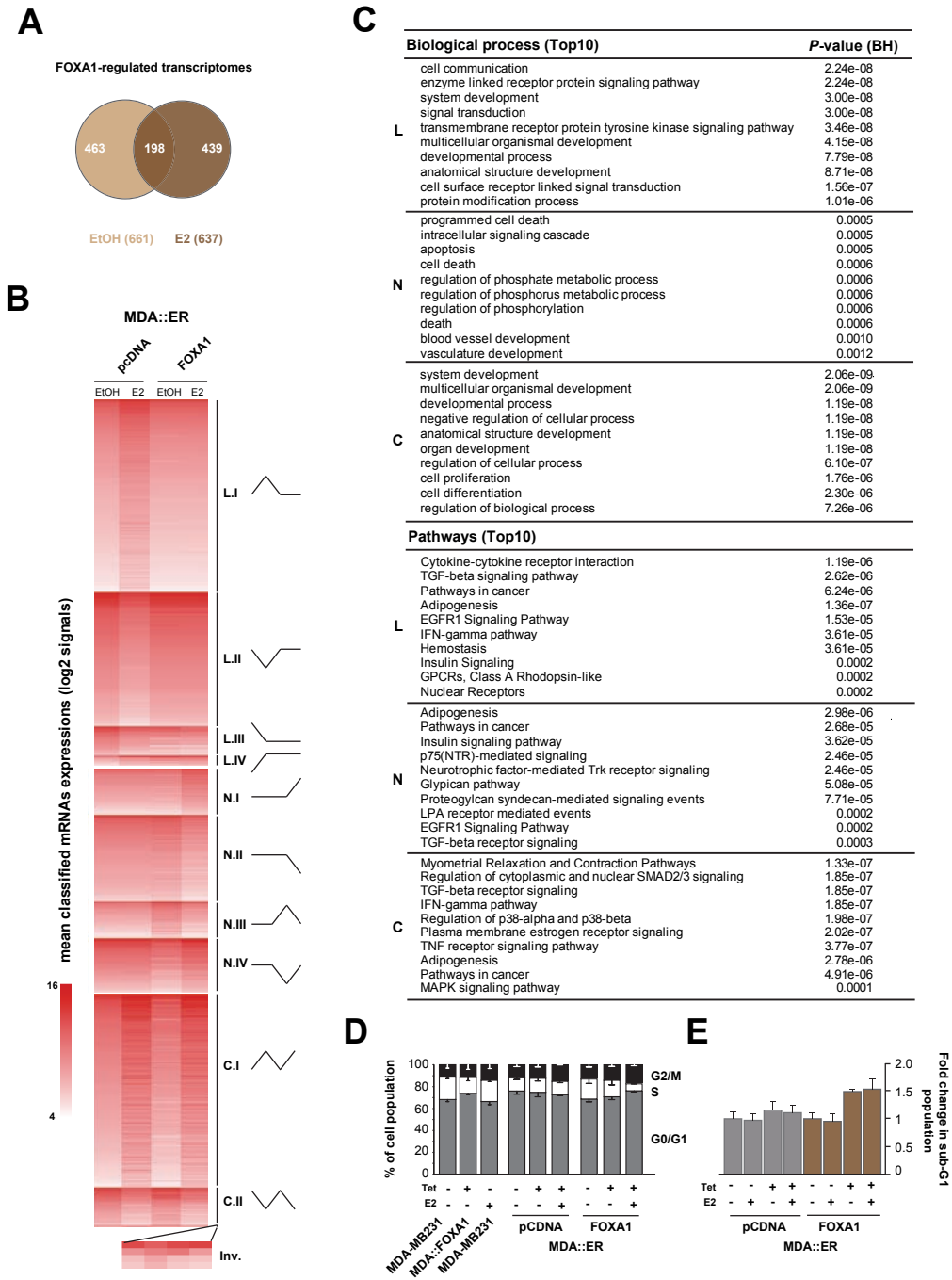


Figure 2

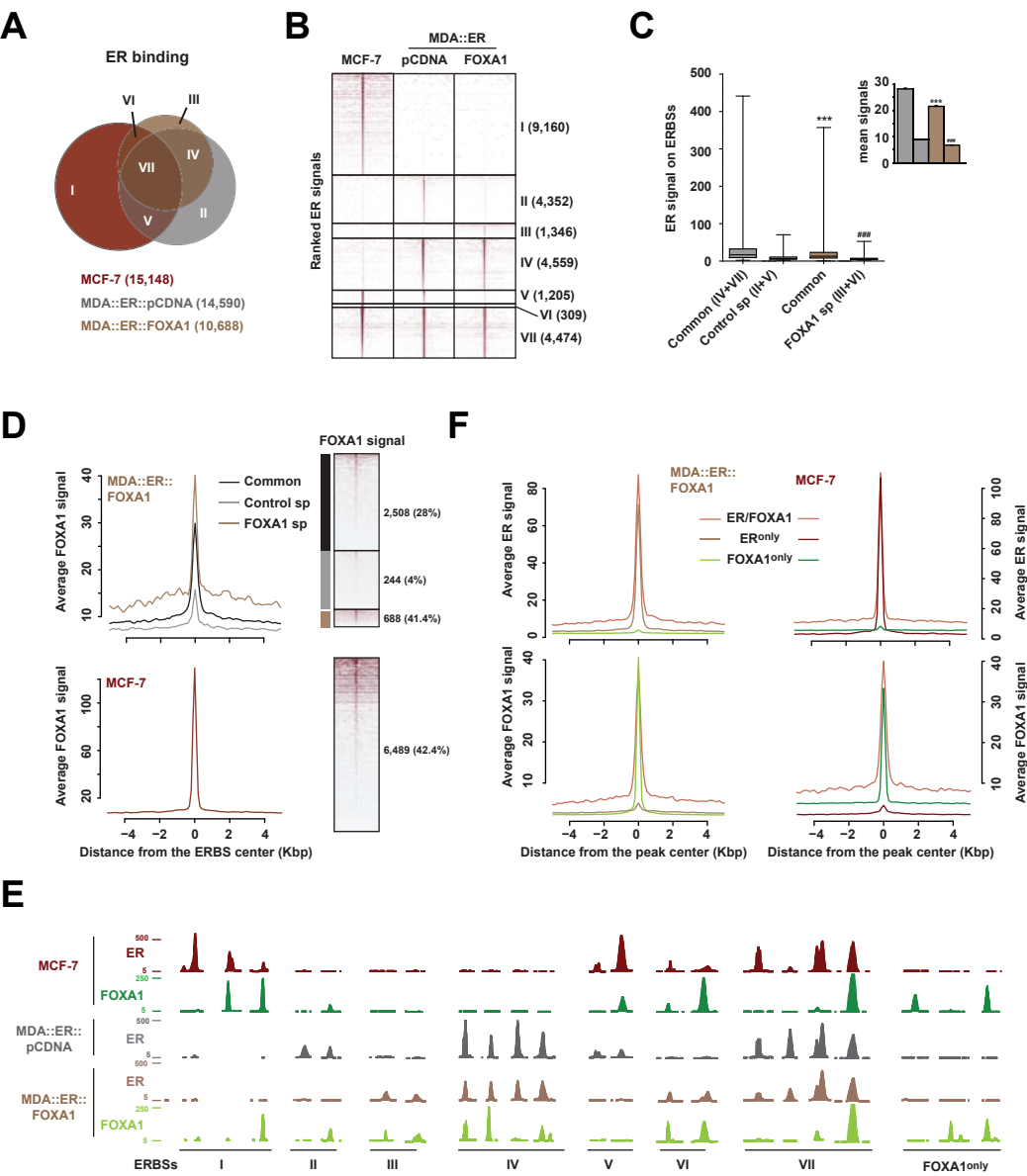


Figure 3

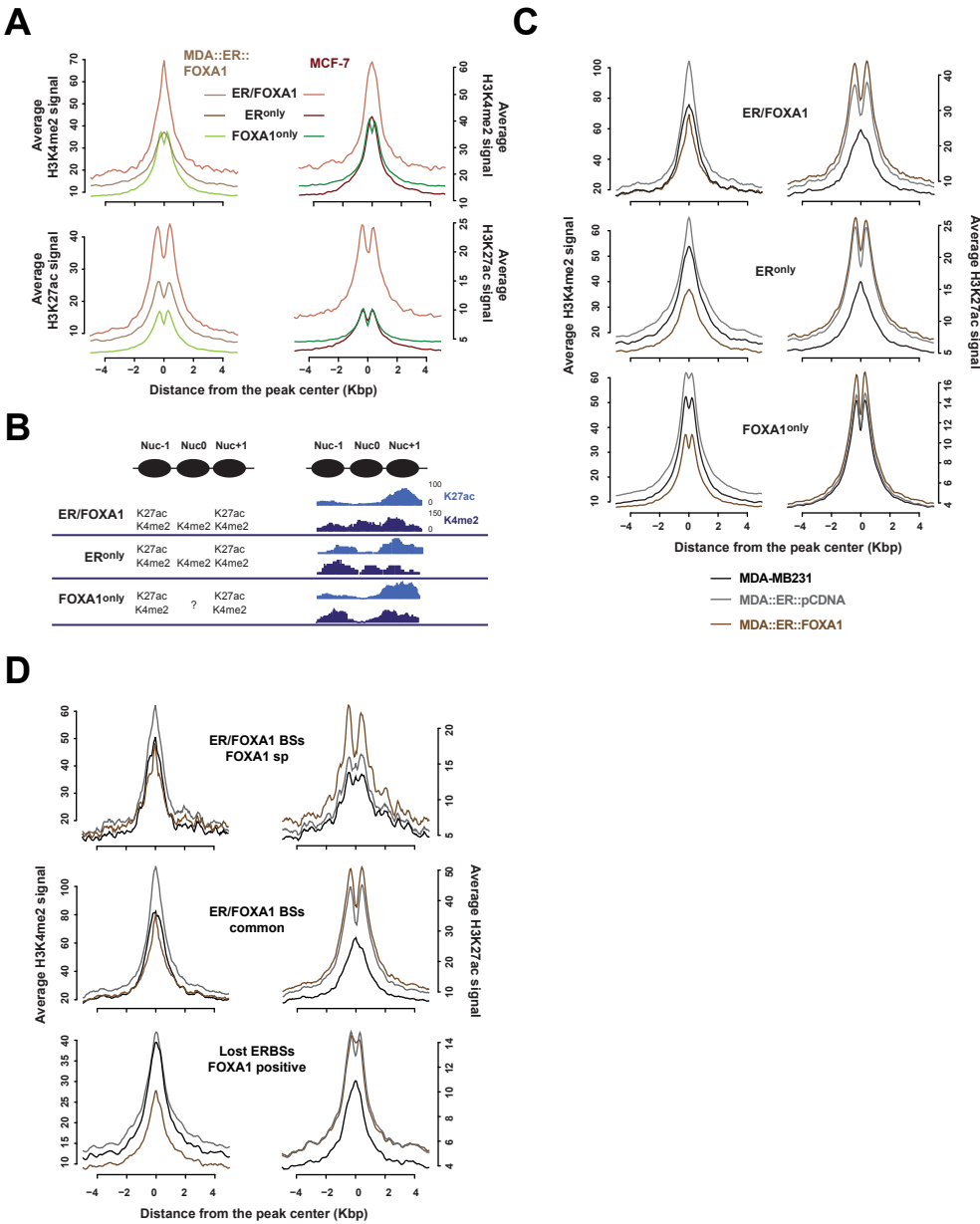


Figure 4

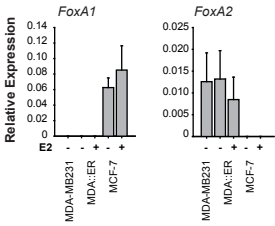
A

MCF-7					
Motif	Rank	P-val	Rank	P-val	Abs [Z-score]
ERE	1	0	1 ESR1		151.72
			2 ESR2		151.50
			3 half-ERE		138.78
AP1	2	2.85E-9	23		48.64
BACH	3	3.4E-8	63		21.07
PAX	4	2.06E-7	44		29.73
CREB	5	4.51E-7	40	1.60E-21	34.18
AP2	6	8.32E-7	85		16.01
MAF	7	1.45E-6	70		19.09
AR	8	1.97E-6	39		34.48
...
FOX	16	0.000576	73		18.14
Centdist			SeqPos		

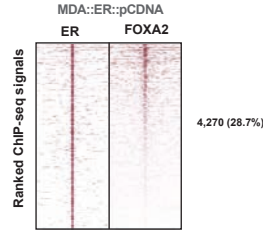
MDA::ER::pCDNA					
Motif	Rank	P-val	Rank	P-val	Abs [Z-score]
ERE	1	0	1 ESR1		149.78
			2 ESR2		148.58
			3 half-ERE		136.50
AP1	2	0	28		49.38
BACH	3	0	52		34.61
MAF	4	1.08E-12	59		29.73
AR	5	3.32E-6	40	1.61E-21	31.00
CREB	6	6.07E-6	46		36.44
PAX	7	1.72E-5	72		23.79
AP2	8	9.05E-5	85		18.82
...
FOX	47	0.00321	132	9.78E-07	5.87
Centdist			SeqPos		

MDA::ER::FOX A1					
Motif	Rank	P-val	Rank	P-val	Abs [Z-score]
ERE	1	0	1 ESR1		125.87
			2 ESR2		124.92
			3 half-ERE		116.83
AP1	2	0	30		40.06
BACH	3	0	52		26.61
MAF	4	5.55E-16	69		21.23
AR	5	2.68E-9	55	1.61E-21	25.91
CREB	6	3.49E-8	23		46.79
PAX	7	4.46E-8	73		18.49
FOX	8	3.96E-7	100	1.34E-18	10.59
SP1	9	9.71E-6	120	9.78E-07	8.84
ETS	10	7.05E-5	97	6.99E-20	10.98
Centdist			SeqPos		

B



C



D

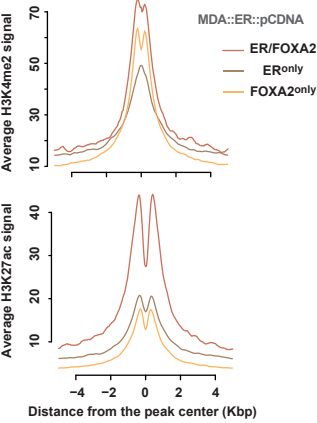


Figure 5

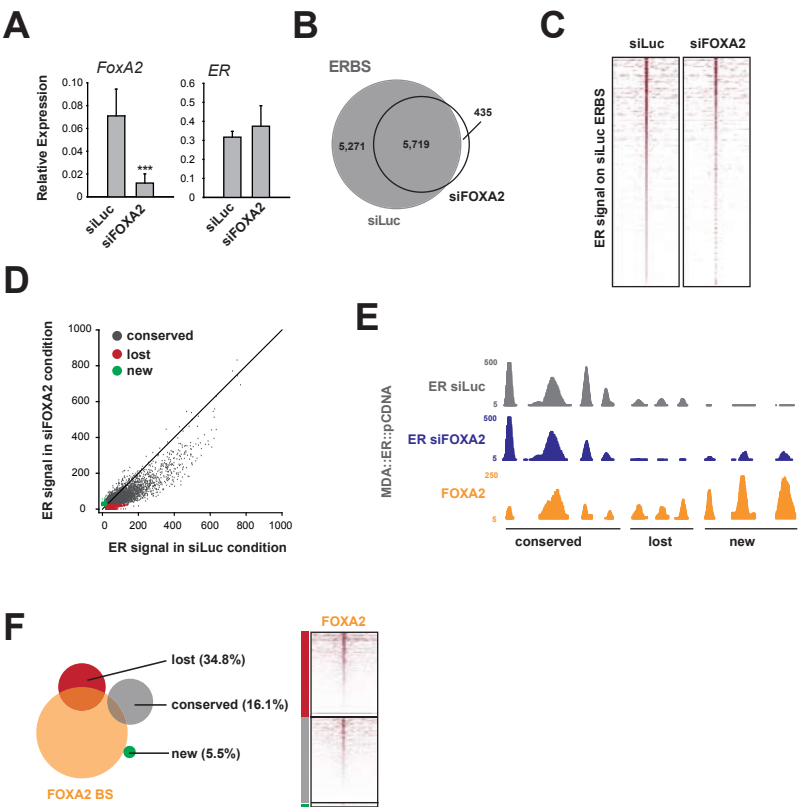


Figure 6

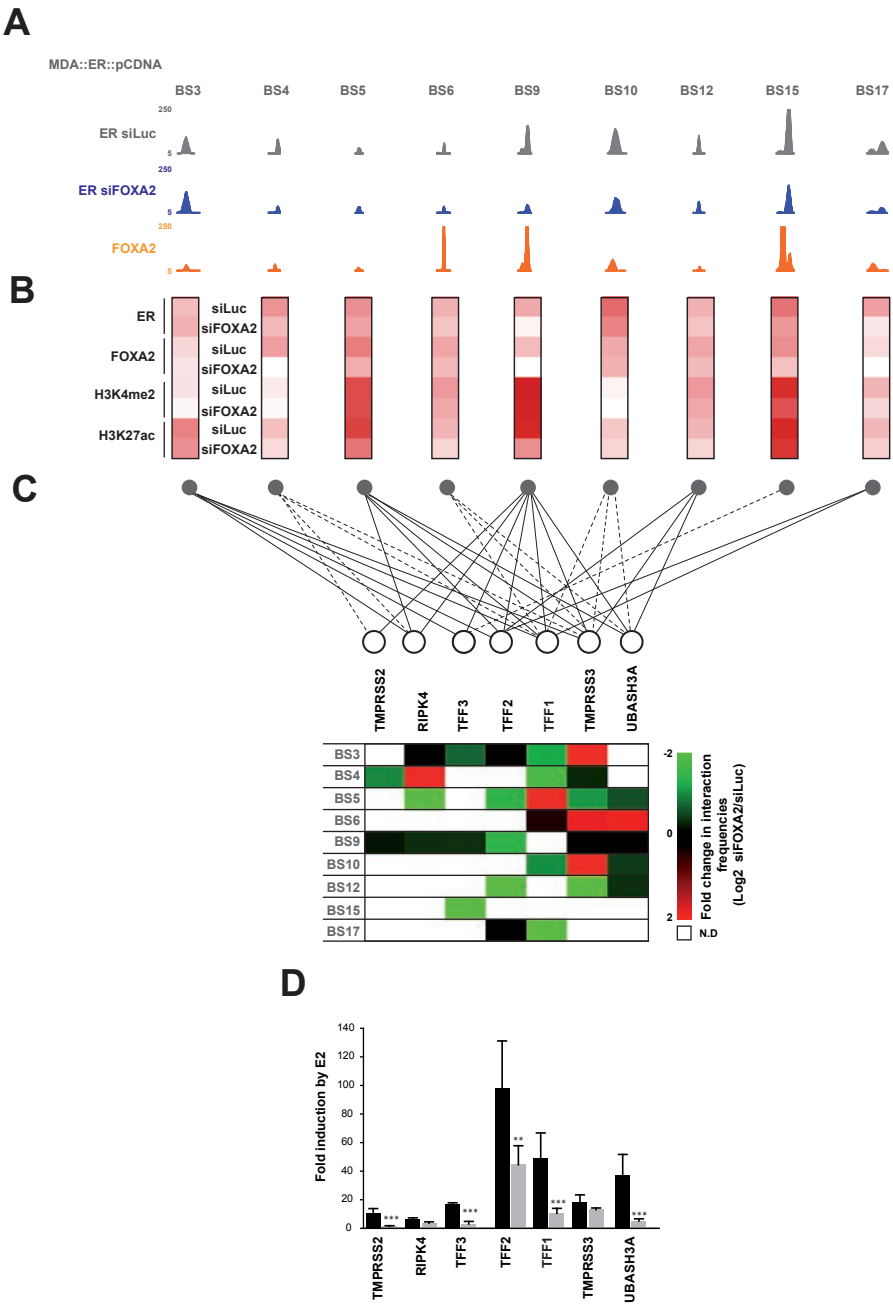
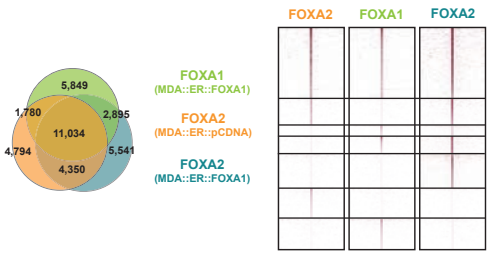
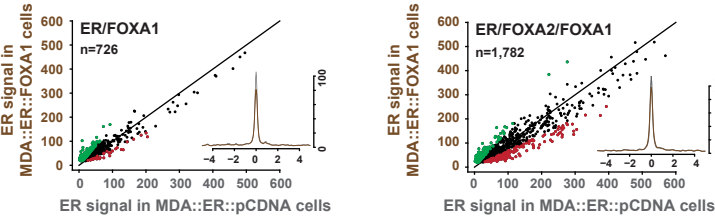


Figure 7

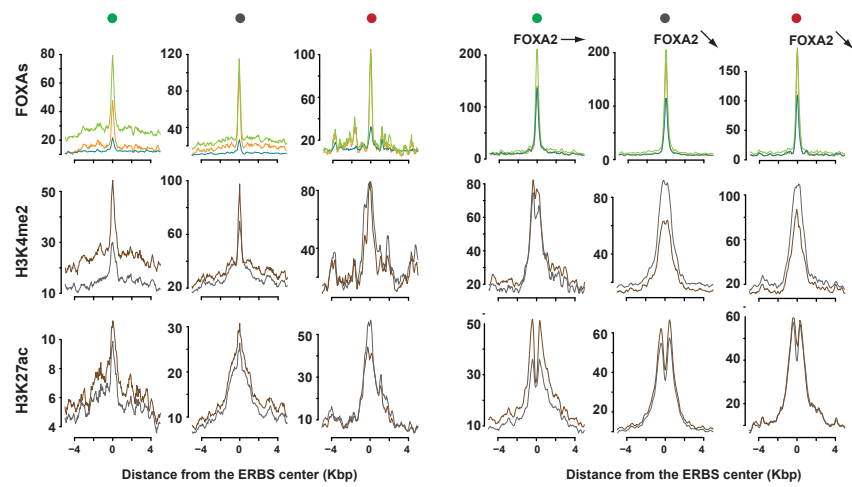
A



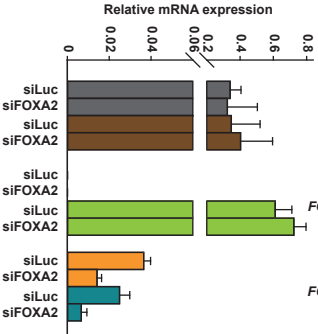
B



C



D



E

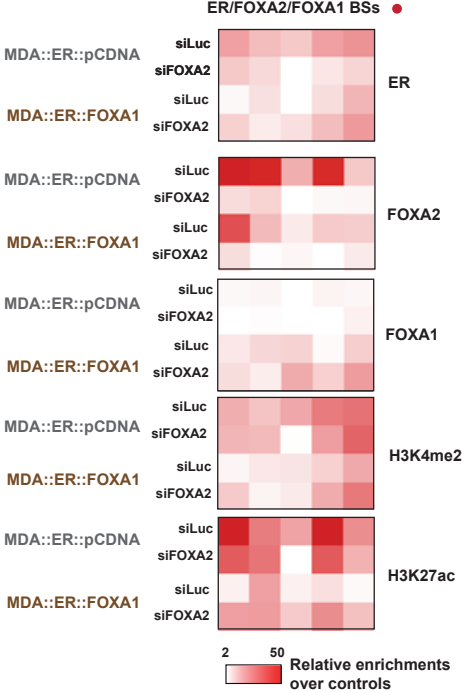


Figure 8

SUPPLEMENTAL TABLES

Table S1. Antibodies used in this study. Multiple references are cited for the same protein when mixes of antibodies were used.

Target	Reference	Provider
ER	HC20	Santa Cruz
ER	ab 10 (TE111-SD11)	NeoMarker
FOXA1	ab 5089	Abcam
FOXA2	clone 3C10	Millipore
FOXA2	ab 23630	Abcam
5mC	MAb-006-100	Diagenode
5hmC	CS-HMC-020	Diagenode
H3	ab 1791	Abcam
H3K4me2	07-030	Millipore
H3K27ac	Ab4729	Abcam

Table S2. ChIP-seq statistics. E4 is the control MDA::ER::pCDNA clone and E8 is the MDA::ER::FOXA1 clone. MDA is the short name for naïve MDA-MB231 cells.

ChIP-Seq libraries	Number of unique mapped reads	Normalized reads	Peak calling Threshold (p-value)	Number of binding sites
MDA::ER Input	23,021,101	NA	NA	NA
MCF7 Input	10,811,755	NA	NA	NA
E4-ER	16,660,633	19,686,939	33 (1e-8)	14,839
E8-ER	19,686,939	19,686,939	28 (1e-8)	10,722
E4-ER-siLuc	19,686,939	19,686,939	21 (1e-6)	12,152
E4-ER-siFOXA2	12,413,498	19,686,939	23 (1e-6)	6,266
MCF7-ER	16,544,063	19,686,939	39 (1e-8)	15,286
E8-FOXA1	19,529,773	57,679,998	60 (1e-8)	21,675
MCF7-FOXA1	19,216,786	57,679,998	56 (1e-8)	27,849
E4-FOXA2	21,036,676	57,679,998	65 (1e-8)	22,079
E8-FOXA2	57,679,998	57,679,998	89 (1e-8)	23,820
E4-H3K4me2	18,069,314	44,812,838	89 (1e-8)	50,549
E8-H3K4me2	11,960,608	44,812,838	51 (1e-8)	66,816
MDA-H3K4me2	13,371,073	44,812,838	76 (1e-8)	61,096
MCF7-H3K4me2	44,812,838	44,812,838	64 (1e-8)	62,836
E4-H3K27ac	17,444,746	17,444,746	48 (1e-8)	26,561
E8-H3K27ac	11,943,804	17,444,746	53 (1e-8)	25,500
MDA-H3K27ac	9,121,253	17,444,746	50 (1e-8)	26,184
MCF7-H3K27ac	10,344,690	17,444,746	43 (1e-8)	23,770

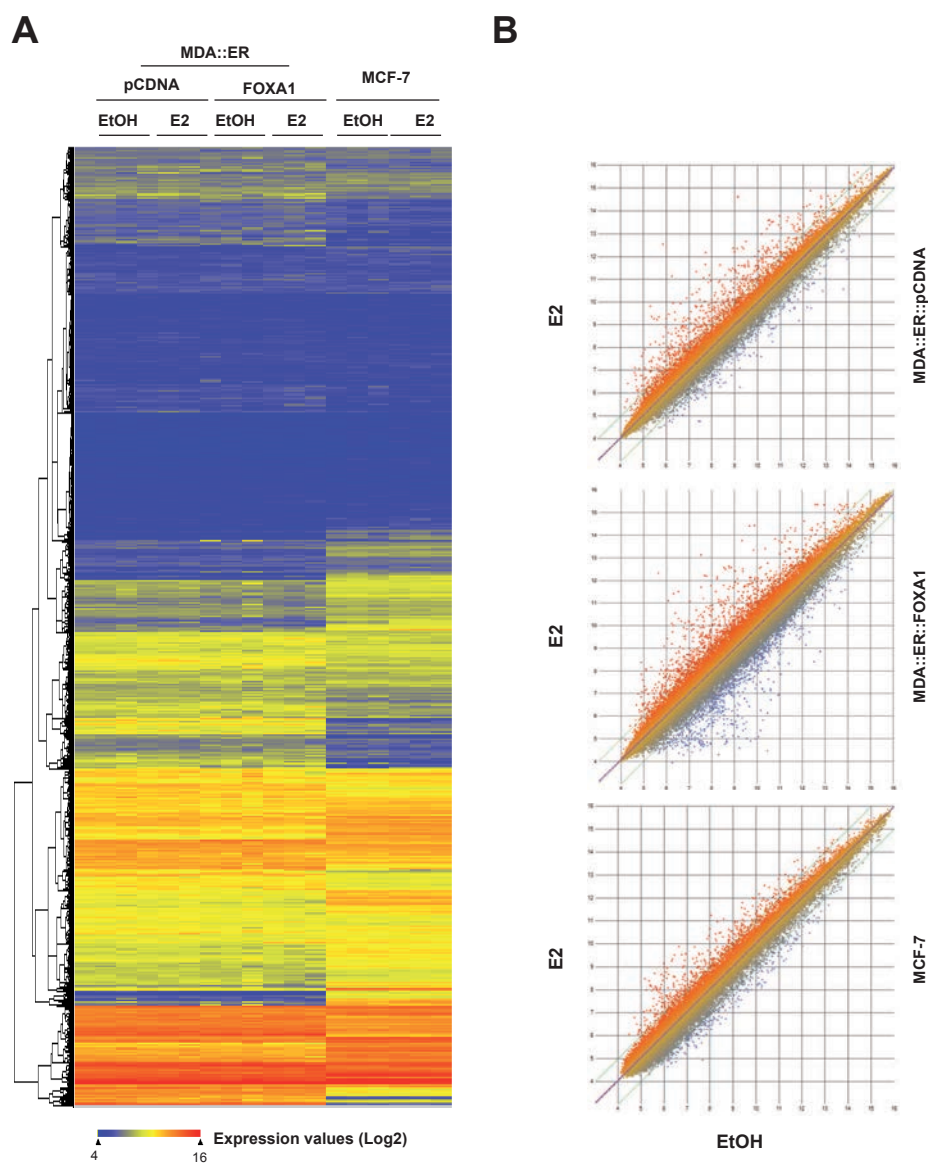


Figure S1. Transcriptomes of MDA::ER::pCDNA, MDA::ER::FOXA1 and MCF-7 cells.

(A) Hierarchical clustering (Euclidian method) of all mRNAs expression values measured in MDA::ER::pCDNA, MDA::ER::FOXA1 and MCF-7 cells for each of the three NimbleGen Homo sapiens 385K Arrays hybridized for each conditions. (B) Scatter plot of mean expression values of all mRNAs detected in estradiol (E2) treated cells vs ethanol (EtOH as vehicle control) treated ones. Graphs were all generated using the ArrayStar software (DNASTar, Inc.).

A

Biological process (Top10)	P-value (BH)
cell proliferation	9.50e-05
cell communication	0.0001
anatomical structure development	0.0001
blood vessel morphogenesis	0.0001
system development	0.0001
DNA replication	0.0001
angiogenesis	0.0002
signal transduction	0.0002
Pathways (Top10)	
Pathways in cancer	1.39e-17
G1 to S cell cycle control	9.05e-10
DNA Replication	9.05e-10
Axon guidance	5.32e-09
Cell cycle	3.06e-08
Metabolic pathways	7.60e-07
Small cell lung cancer	8.51e-07
Focal adhesion	9.86e-07
DNA damage response	2.32e-06
p53 signaling pathway	9.84e-06

B

Biological process (Top10)	P-value (BH)
cell migration	3.82e-07
localization of cell	1.54e-06
cell motility	1.54e-06
cellular component movement	1.88e-06
locomotion	4.95e-06
response to external stimulus	2.80e-05
response to stress	3.58e-05
wound healing	3.58e-05
response to wounding	4.08e-05
muscle cell proliferation	0.0001
Pathways (Top10)	
Pathways in cancer	4.91e-11
DNA Replication	2.20e-08
Complement and Coagulation Cascades	8.50e-08
Adipogenesis	1.84e-07
Axon guidance	1.78e-06
Cell adhesion molecules	1.78e-06
MAPK signaling pathway	1.78e-06
Senescence and Autophagy	3.11e-06
Cytokine-cytokine receptor interaction	6.86e-06
TGF Beta Signaling Pathway	9.92e-06

Figure S2. Functional annotations of genes whose expression is affected by FOXA1.

Webgestal services (hosted at www.bioinfo.vanderbilt.edu/webgestalt/; Duncan et al., 2010) were used to functionally annotate genes regulated by FOXA1 (MDA::ER::FOXA1 vs MDA::ER::pCDNA cells). Top ten GO terms and pathways (compilations of Kegg, common and wiki pathways) are indicated with associated Benjamini&Hochberg (BH) corrected P-value for genes significantly regulated by FOXA1 in the absence (A) or presence of E2 (B).

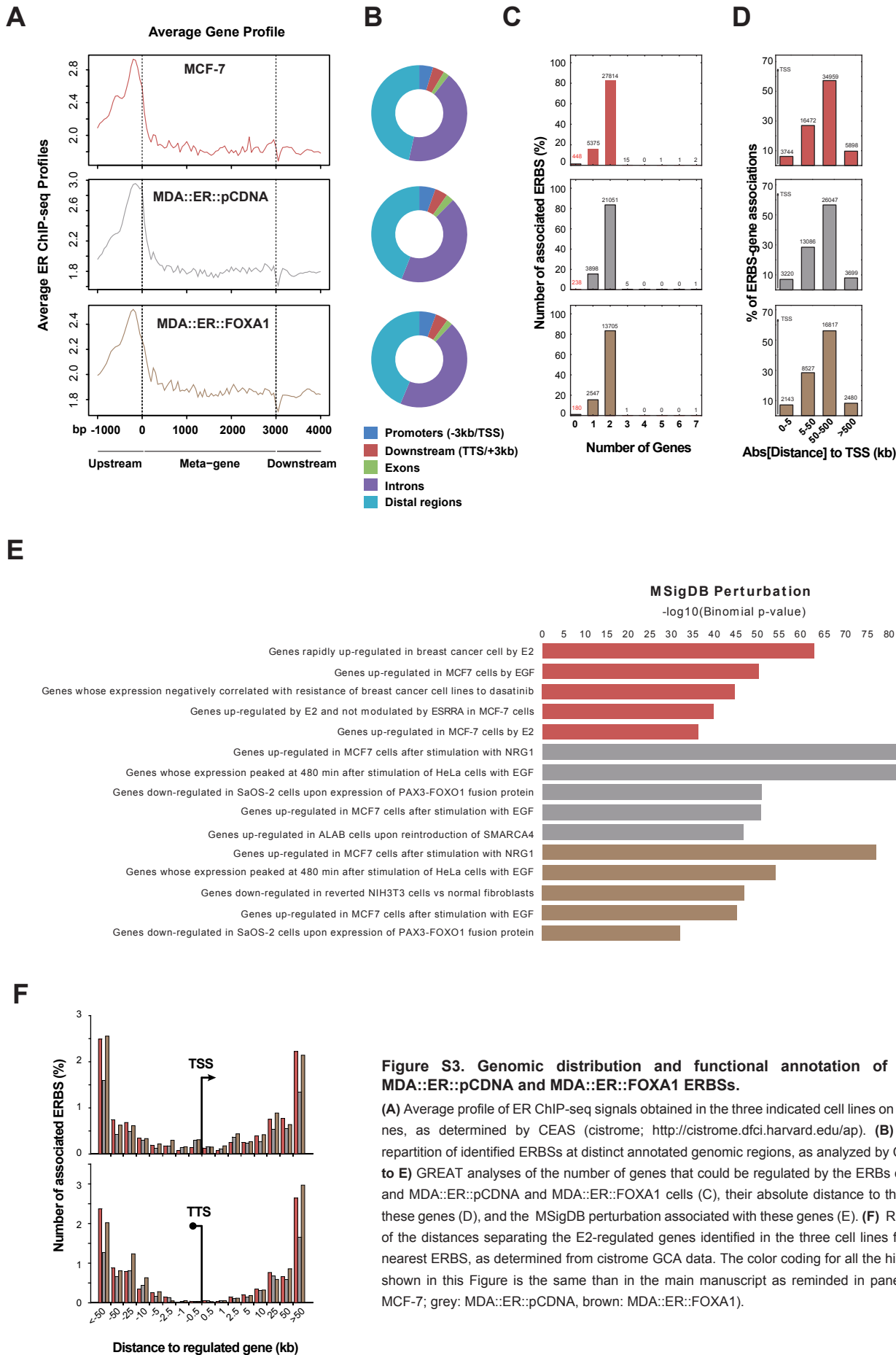


Figure S3. Genomic distribution and functional annotation of MCF-7, MDA::ER::pCDNA and MDA::ER::FOXA1 ERBSs.

(A) Average profile of ER ChIP-seq signals obtained in the three indicated cell lines on meta-genes, as determined by CEAS (cistrome; <http://cistrome.dfci.harvard.edu/ap>). (B) Relative repartition of identified ERBSs at distinct annotated genomic regions, as analyzed by CEAS. (C to E) GREAT analyses of the number of genes that could be regulated by the ERBS of MCF-7 and MDA::ER::pCDNA and MDA::ER::FOXA1 cells (C), their absolute distance to the TSS of these genes (D), and the MSigDB perturbation associated with these genes (E). (F) Repartition of the distances separating the E2-regulated genes identified in the three cell lines from their nearest ERBS, as determined from cistrome GCA data. The color coding for all the histograms shown in this Figure is the same than in the main manuscript as reminded in panel A (red: MCF-7; grey: MDA::ER::pCDNA, brown: MDA::ER::FOXA1).

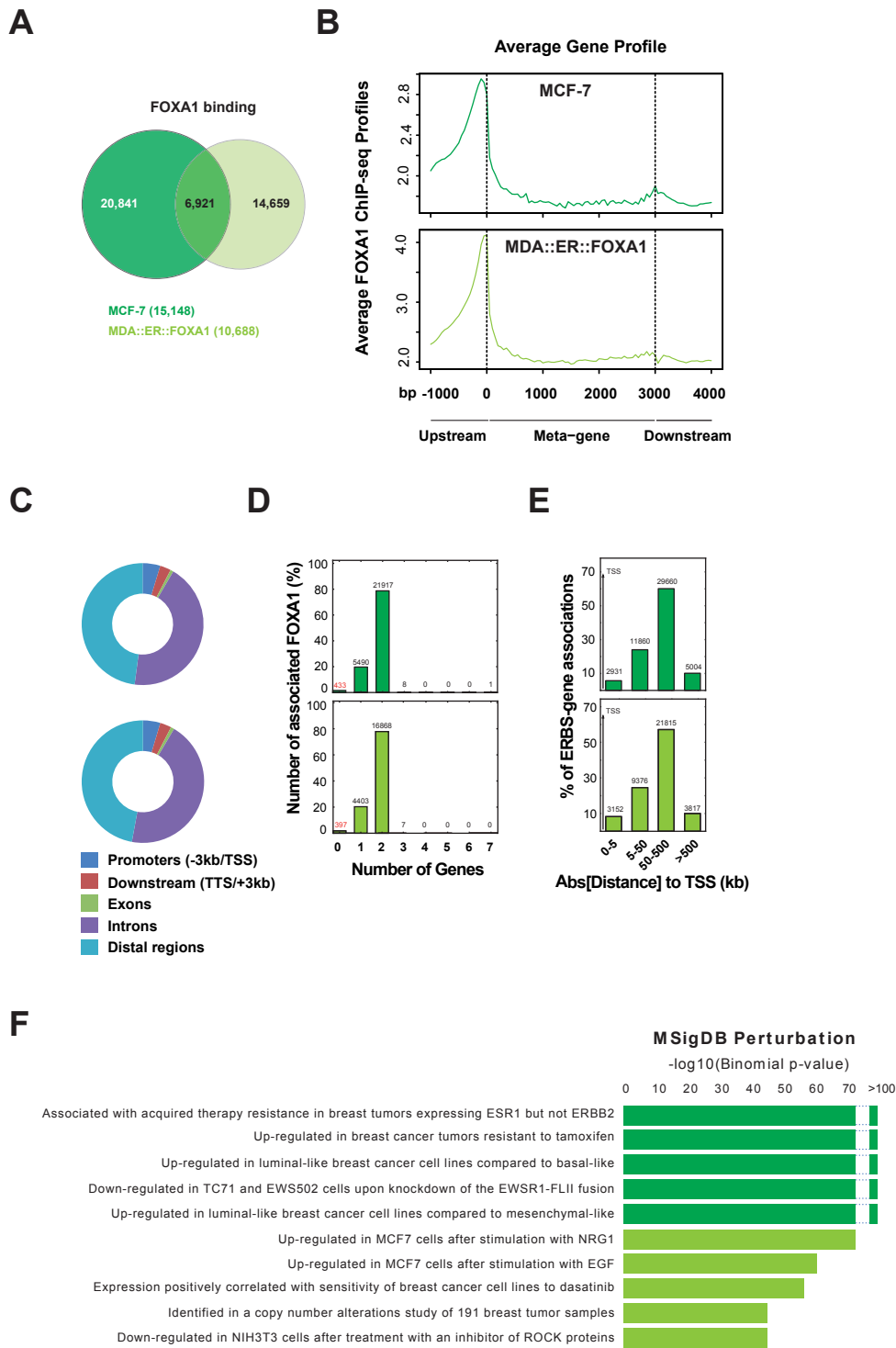


Figure S4. Genomic distribution and functional annotation of MCF-7 and MDA::ER::FOXA1 FOXA1BSs.

(A) Overlap of FOXA1 BSs between MCF-7 and MDA::ER::FOXA1 cells. (B) Average profile of FOXA1 ChIP-seq signals obtained in the two indicated cell lines on meta-genes, as determined by CEAS (cistrome). (C) Relative repartition of identified FOXA1BSs at distinct annotated genomic regions, as analyzed by CEAS. (D to E) GREAT analyses of the number of genes that could be regulated by the FOXA1BSs of MCF-7 and MDA::ER::FOXA1 cells (D), their absolute distance to the TSS of these genes (E), and the MSigDB perturbation associated with these genes (F). The color coding for all the histograms shown is the same than in the main manuscript as reminded in panel B (green: MCF-7; light green: MDA::ER::FOXA1).

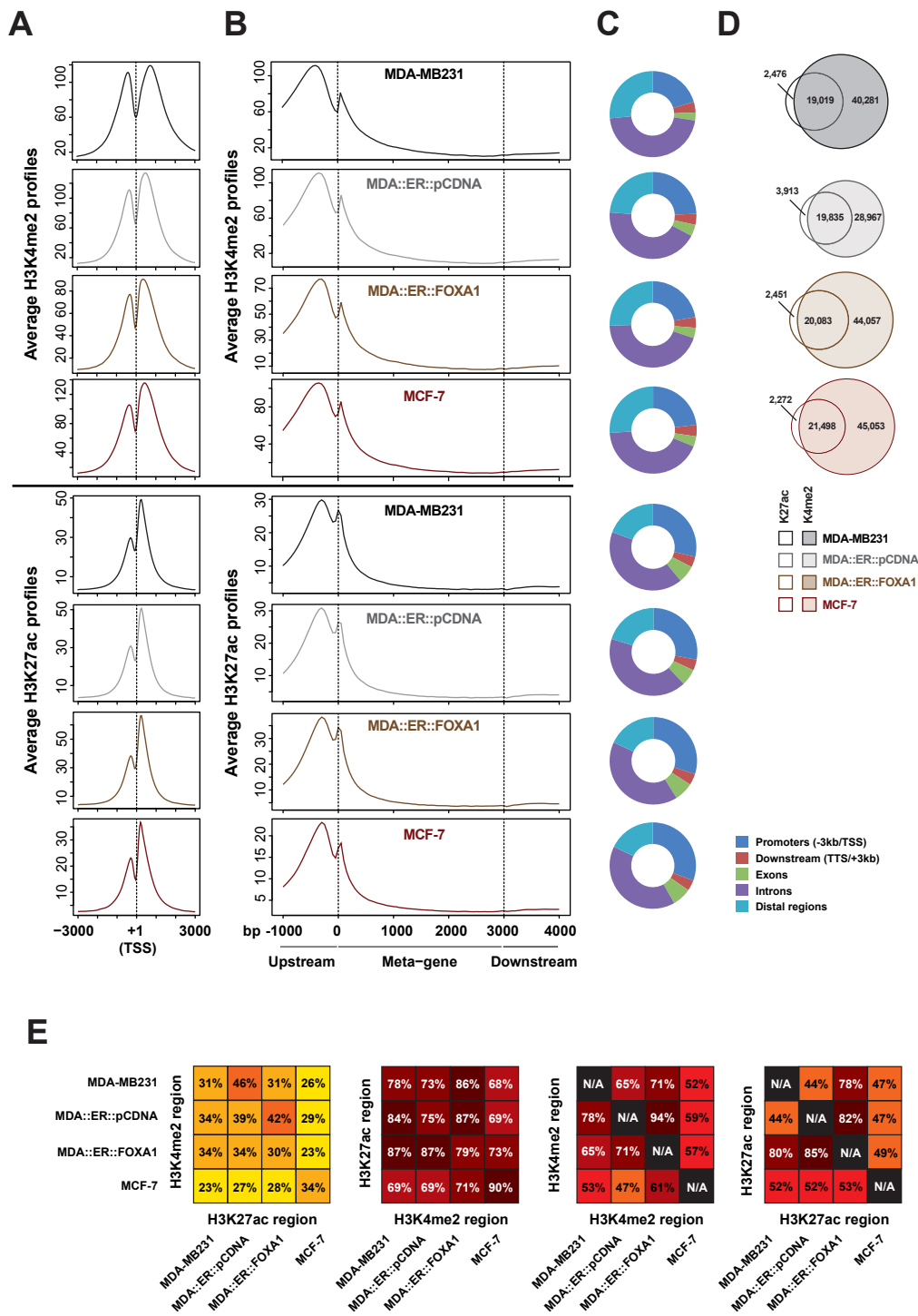


Figure S5. Genomic distribution and annotations of H3K27ac and H3K4me2 signals in all studied cell lines.

(A) Average profile of H3K4me2 and H3K27ac ChIP-seq signals obtained in indicated cell lines in a -3/+3 kb window centered on the transcription start site (+1, TSS), or on meta-genes (B), as determined by CEAS (cistrome). (C) Relative repartition of identified genomic regions enriched for either histone post-translational modifications at distinct annotated genomic regions, as analyzed by CEAS. (D) Overlap of regions enriched in H3K4me2 and H3K27ac marks for each considered cell lines. (E) Combination matrices illustrating the relative overlap of regions enriched for either histone marks in the 4 considered cell lines.

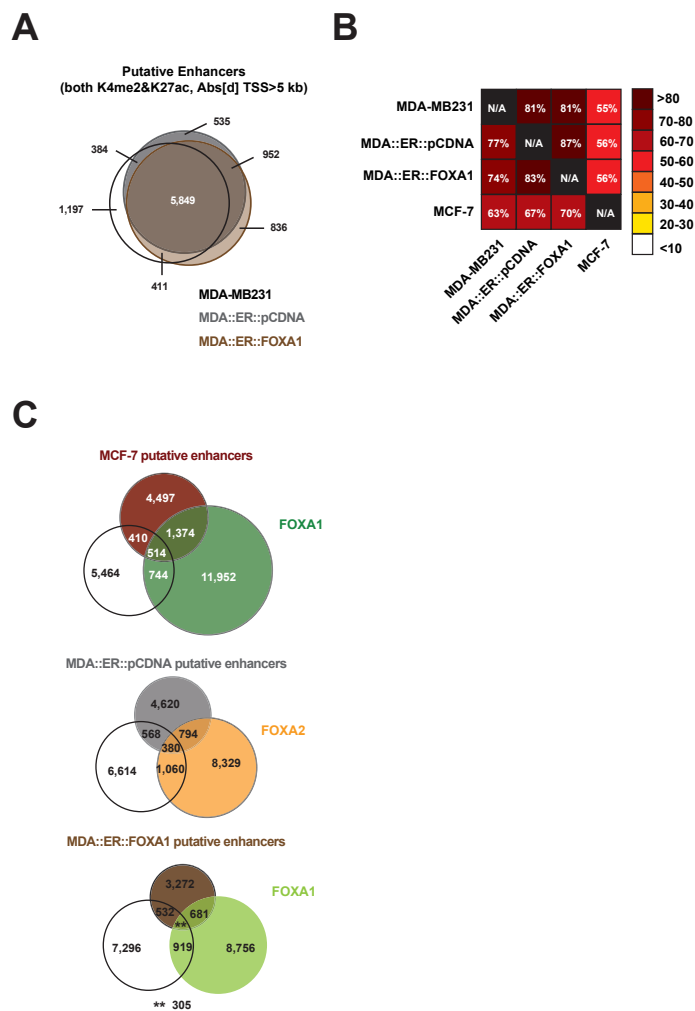


Figure S6. Common H3K27ac and H3K4me2 positive regions signals in all four cell lines.
(A and B) Overlap of H3K27ac/K4me2 positive regions distant of at least 5 kb from any genes TSSs in MDA-MB231, MDA::ER::pCDNA and MDA::ER::FOXA1 cells depicted as a Venn diagram (A) or as a matrix (B) including MCF-7 sites. (C) Venn diagrams depicting the overlap between the FOXA1 or FOXA2 sites located at least 5 kb away from genes TSSs with these putative enhancers.

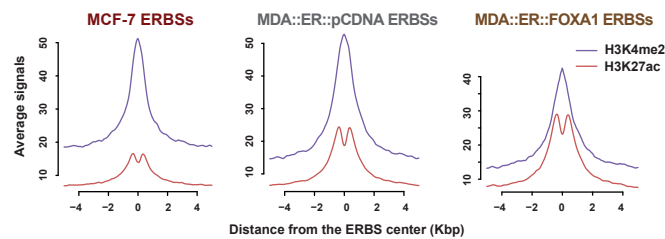
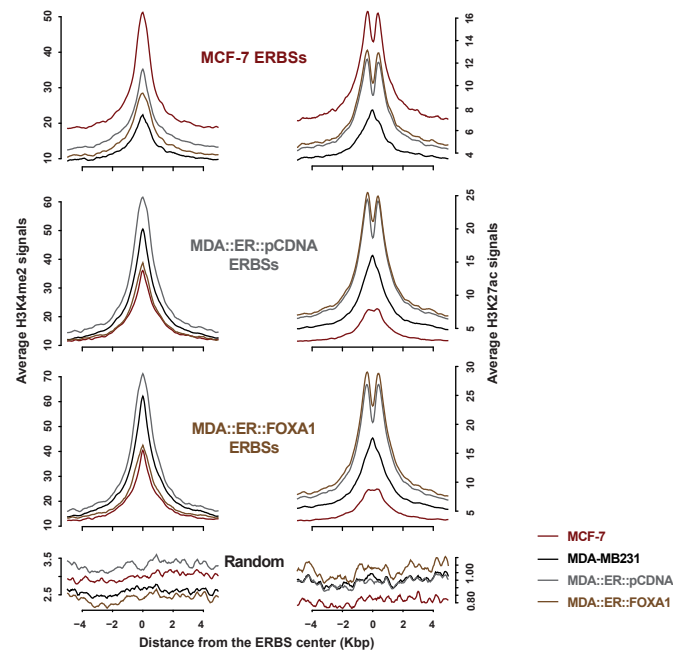
A**B**

Figure S7. Genomic distribution of H3K27ac and H3K4me2 signals on ERBSs.

(A) Average profiles of H3K4me2 and H3K27ac ChIP-seq signals obtained in each of the three cell lines on a -5/+5 kb window surrounding their genomic regions mobilizing ER. (B) Mean intensities of H3K4me2 (left side) and H3K27ac (right side) enrichments in each of the indicated cell lines within a -5/+5 kb window centered on ERBSs determined in each cell lines. As a control, the same calculations were made using 25 sets of random regions (containing 25,000 genomic fragments selected for a mean length identical as ER positive regions). One of these control profiles is shown at the bottom of the panel.

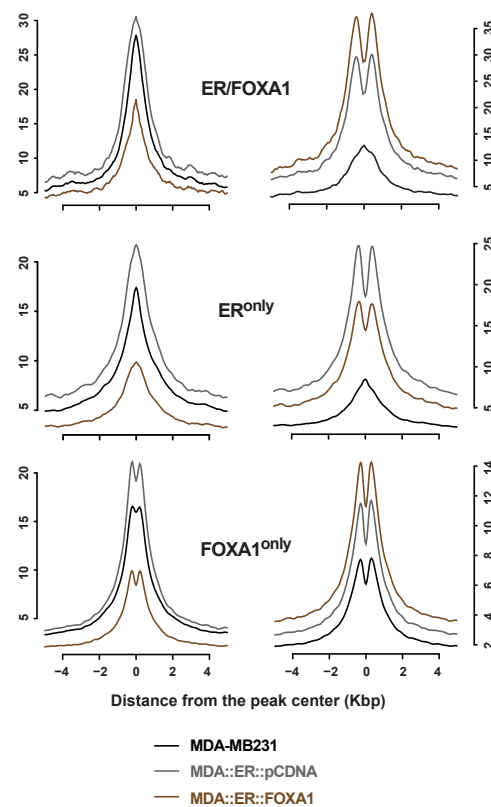


Figure S8. Genomic distribution of non-normalized H3K27ac and H3K4me2 signals.

Average profiles of H3K4me2 and H3K27ac ChIP-seq signals obtained in each of the three indicated cell lines on a -5/+5 kb window surrounding their genomic regions mobilizing ER. In contrast to figures from the main manuscript, we used here signals which were not normalized by their relative sequencing depth. Conclusions made by both approaches (normalized vs. not) are similar.

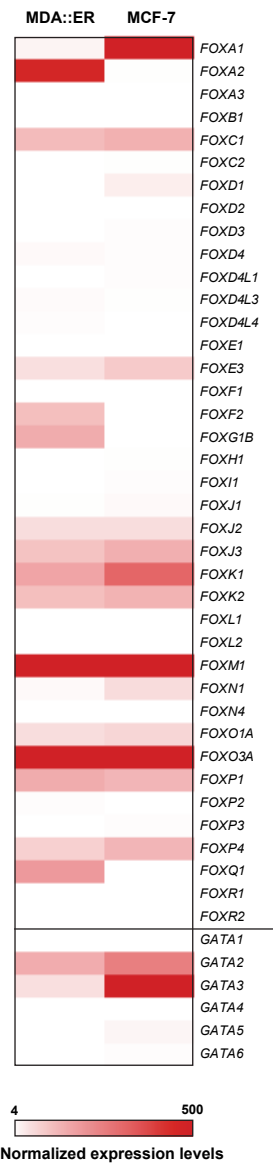


Figure S9. Expression of Forkhead and GATA factors in MDA::ER and MCF-7 cells.

This heatmap represents expression values detected in MCF-7 and MDA::ER cells for each of the mRNAs annotated on the right side as taken from our transcriptomic data (GSE 23850).

DISCUSSION

L'ensemble du travail présenté dans la première partie des résultats nous a permis de définir finement les mécanismes dynamiques à l'origine des régulations transcriptionnelles coordonnées du cluster de gènes œstrogéno-dépendant TFF ; ceci à la fois dans les cellules MCF-7 qui expriment ER de façon endogène ainsi que dans les cellules MDA::ER qui l'expriment de façon ectopique. L'utilisation de ces deux lignées en parallèle nous a permis de montrer que les mécanismes mis en jeu dans la régulation de ce cluster étaient, de par leur nature, globalement similaires; suggérant donc que l'expression d'une protéine ER dans un contexte initialement neutre en terme de réponse œstrogénique était capable par elle seule de mettre en place et de reproduire des processus de régulation identiques à ceux se produisant dans un contexte relativement plus physiologique. Cette observation reste cependant à tempérer par le fait que, comme décrit par d'autres auteurs dans ce type de lignées cancéreuses mammaires (Garcia *et al.*, 1992) ou provenant d'autres types cellulaires comme les U2OS (Krum *et al.*, 2008), les transcriptomes œstrogéno-dépendants, ainsi que les cistromes ER ne sont que partiellement ressemblants. Ainsi, si l'on se limite au cluster TFF, il apparaît que le nombre de gènes régulés par les œstrogènes ainsi que le nombre de sites de liaison du ER diffèrent entre les cellules MDA::ER et les MCF-7. De même pour le nombre et la nature des interactions entre séquences régulatrices et promoteurs de gènes régulés.

Un des points les plus importants de cette étude a reposé sur la mise au point d'un nouvel outil permettant d'analyser les mécanismes initiés par l'ER au niveau de ses séquences cibles sur la chromatine: les TFOs (Triplex Forming Oligonucleotide). L'utilisation de ces petits oligonucléotides nous a en effet permis d'inhiber spécifiquement le recrutement d'ER au niveau de régions précises du génome, et donc de caractériser quelles étaient les contributions exactes de la mobilisation du ER sur ces séquences pour la régulation d'un gène donné. Une des conclusions majeures issues de ces expériences est la confirmation d'une hypothèse de redondance fonctionnelle entre enhancers qui avait été posée dès les premières caractérisations des cistromes d'ER, que ce soit à l'échelle de quelques chromosomes au départ (Carroll *et al.*, 2005; Hurtado *et al.*, 2011) ou à l'échelle du génome. Celle-ci se basait sur l'observation que les sites de recrutement d'ER avaient une certaine tendance à se retrouver organisés en clusters de sites, dispersés dans le génome: les sites de liaison de ER sont en effet rarement des sites isolés dans le génome. L'utilisation des TFOs nous a permis de montrer, à l'échelle du cluster TFF, que la régulation d'un gène par ER pouvait en effet impliquer des contacts entre plusieurs enhancers et le promoteur de ce gène, et que l'abolition

du recrutement d'ER sur une de ces régions pouvait être contrebalancée par les processus mis en place par ER au niveau des autres enhancers.

Il serait bien sur extrêmement intéressant de pouvoir étendre ce type d'étude à de plus grandes échelles, voire même à l'échelle du génome. Ceci pourrait en effet permettre de fonctionnaliser précisément les cistromes et interactomes d'ER en répondant à cette question simple qui pourtant reste encore relativement obscure malgré l'avènement des techniques à grande échelle: qui régule quoi? Cependant cette technique présente deux limites majeures. La première repose sur la difficulté à concevoir des TFOs répondant à tous les critères de sélection nécessaires à l'établissement de structures d'ADN en triplex, notamment en terme de composition des séquences cibles. Ainsi, en diminuant la stringence aux limites des caractéristiques modélisées par Vekhoff *et al.* (2008) , 20% des sites de liaison d'ER au maximum (MCF-7 ou MDA::ER) pourraient effectivement être ciblés par ce type d'approche. Sachant que sur ces 20%, le nombre de TFOs ayant une réelle efficacité serait de toute façon inférieur. Nous avons par exemple testé plus d'une quarantaine d'oligonucléotides avant d'en retenir une dizaine comme étant capables de i) se lier aux séquences cibles et ii) de réduire ou d'affecter la mobilisation d'ER. Il n'empêche que cette technique est dans l'utilisation beaucoup plus simple que ce qu'il faudrait mettre en œuvre lors d'utilisation de modificateurs tels que les ZFNs ("Zinc Finger Nucleases") ou les TALENs ("Transcription Activator-Like Effector Nuclease") permettant de générer des modifications génétiques ponctuelles (Chen *et al.*, 2011; Carlson *et al.*, 2012). L'utilisation possible de TFOs permettrait donc, dans certains cas, de définir quels sont les enhancers préférentiels d'un gène, ou inversement d'identifier quels gènes sont contrôlés par un enhancer particulier. La seconde limite des TFOs concerne ce dernier point: une très forte redondance entre enhancers pourrait masquer et donc biaiser les conclusions que des analyses à grande échelle (études transcriptomiques par exemple) pourraient générer.

Un des apports que pourraient donner les TFOs dans la compréhension des interactomes entre sites de liaison d'ER et promoteurs de gènes régulés ou entre les sites ER entre eux serait un éclaircissement des nombre et importance fonctionnelle des boucles chromatinienne inter-chromosomiques. En effet, si les données de ChIA-PET ER n'ont pas permis d'identifier de telles boucles (Fullwood *et al.*, 2009b), les techniques de 4C et de Hi-C ont au contraire mis en avant l'existence d'un nombre relativement important de contacts inter-chromosomiques (Sanyal *et al.*, 2012). Leur nature exacte (boucles cohesine-dépendantes ou bien simple proximité dans l'espace par exemple) reste pour le moment peu caractérisée,

même si les récentes données de ChIA-PET CTCF ont dans ce cas permis d'identifier de tels phénomènes (Handoko *et al.*, 2011). Par exemple, une boucle de chromatine entre le site ER situé au sein du promoteur du gène *TFF1* et le promoteur du gène *GREB1* situé sur le chromosome 2 a été identifiée par 4C (Hu *et al.*, 2008). Nous avons pu réaliser une hybridation pilote de nos échantillons de 4C préparés à partir de ce site de liaison ainsi que de l'ERBS 2 du cluster TFF dans les cellules MDA::ER; et, de façon intéressante, un pic a été détecté au niveau du promoteur du gène *GREB1*. De plus, la transfection des TFOs correspondant s'est avérée avoir un impact négatif sur la régulation œstrogénique de ce gène. Un rapprochement de ces séquences dans l'espace nucléaire semble donc être possible, et posséder une importance fonctionnelle. Cependant, nos essais de caractérisation de cette interaction dans les cellules MDA::ER par DNA FISH ont tous été infructueux, comme montré dans une autre étude dans les cellules MCF-7 (Kocanova *et al.*, 2010). Se pose donc la question de savoir quelles sont les limites exactes de ces différentes techniques d'analyse afin de caractériser les interactions inter- voire intra-chromosomiques: DNA FISH vs 3C et dérivés. Par exemple, en DNA-FISH, quel serait le seuil à appliquer pour parler de proximité physique entre deux régions génomiques, de par la nature potentiellement stochastique de ces événements de rapprochements ? Dans quelle limite les deux sondes doivent-elles colocaliser en terme de fréquence de détection et en terme de surface de recouvrement ? Il semblerait donc que les techniques de 3C/4C complémentées d'une caractérisation fonctionnelle des interactions détectées par des TFOs pourrait être une combinaison appropriée pour définir l'existence de boucles de chromatine. L'utilisation de TFO, bien que potentiellement réduite dans son efficacité par la redondance des enhancers, pourrait en effet lever un des problèmes d'interprétation des données de 3C/4C lié à ces techniques en elles-mêmes qui utilisent des étapes d'amplification sur du matériel provenant de populations cellulaires et qui peuvent donc détecter des événements rares et possiblement stochastiques. Par exemple, si une interaction détectée n'a effectivement lieu que dans une faible proportion de cellules à un temps t , un TFO pourrait ou non valider son importance fonctionnelle.

Il semble apparaître que malgré les différents cistromes et interactomes caractérisés, la notion de boucles de chromatine reste pour le moment relativement vague en ce qui concerne leurs points d'ancrage. S'il est clair de par nos résultats de 3C et de DNA-FISH que CTCF et la cohésine peuvent effectivement jouer un rôle dans l'organisation de la structure tridimensionnelle du cluster TFF, il serait intéressant d'essayer de caractériser à l'aide de TFOs spécifiques lequel ou lesquels des sites de liaison de ces protéines sont effectivement à

l'origine de boucles (si elles existent réellement *stricto sensu*) organisant cette région génomique. Ceci permettrait peut-être d'aboutir à une compréhension plus fine de l'organisation du cluster TFF comme elle a été obtenue pour le cluster *H19/IGF2* (Nativio *et al.*, 2008). De plus, nos résultats de ChIP-chip montrent que si les cistromes CTCF des cellules MCF-7 et MDA::ER sont extrêmement similaires, ce n'est pas le cas pour la cohésine (une observation que nous avons depuis confirmée à l'échelle du génome par ChIP-seq). Ceci suggère que l'organisation même du cluster TFF diffère entre les deux types cellulaires, avec des boucles de chromatine possiblement différentes pouvant donc avoir une conséquence sur l'interactome reliant les sites de liaison de ER et les promoteurs de ces gènes cibles.

Un autre point important de ce premier travail a été la détermination de l'importance des facteurs pionniers dans les mécanismes dynamiques et coordonnés de régulation de l'activité du cluster TFF, et particulièrement celle des facteurs de la famille GATA: GATA3 dans le cas des cellules MCF-7 et GATA2/3 pour les MDA::ER. Le rôle des facteurs GATAs dans la signalisation œstrogénique était une notion déjà émergente, GATA3 ayant été déterminé comme faisant partie d'un réseau de régulations œstrogénique (Eeckhoute *et al.*, 2007; Wilson et Giguere, 2008). De plus, GATA3 est un des facteurs clefs dans le développement de la glande mammaire (Kouros-Mehr *et al.*, 2006; Asselin-Labat *et al.*, 2007). Cependant les processus de régulations des activité du ER par cette protéine n'étaient pas connus. Le mécanisme d'action de ces facteurs que nous avons mis en évidence au niveau du cluster TFF est relativement original, car il ne rentre pas dans le cadre de ceux mis en évidence dans le cas d'autres facteurs pionniers; tels FOXA1 qui est le pionnier d'ER le plus décrit (Carroll *et al.*, 2006 ; Hurtado *et al.*, 2011), ou encore PBX1 décrit plus récemment (Magnani *et al.*, 2011). En effet, contrairement à ces protéines, l'action de GATA3 ou GATA2 dans les MDA::ER semble être de limiter l'action de FOXA1 et ER en modulant l'intensité et l'étendue de modifications de l'ADN, en particulier de H3K4me2. Et ceci à des étapes apparemment précoces lors de la fonctionnalisation des enhancers du cluster TFF puisque ces processus semblent se dérouler en absence d'E2. Il faut noter que depuis l'accomplissement de ce travail, le laboratoire de J. Carroll a caractérisé ce type de mécanisme limitant à l'échelle du génome dans les cellules MCF-7 (Theodorou *et al.*, 2013) sur une sous-population des sites ER/FOXA1/GATA3.

Le second point abordé lors de ma thèse, faisant suite à ces problématiques d'action pionnière d'autres facteurs de transcription sur l'activité transcriptionnelle du ER, s'est

concentrée sur les facteurs de la famille forkhead. En effet, nous avons postulé à l'issue des travaux présentés dans le premier manuscrit qu'un facteur non caractérisé pouvait manquer au sein des MDA::ER afin de reproduire parfaitement les effets exercés par GATA3 au sein des cellules MCF-7. D'autre part, le faible recouvrement en terme de cistrome ER entre ces deux lignées et le fait que les cellules MDA::ER n'expriment pas FOXA1, nous avaient fait émettre l'hypothèse que ce facteur manquant pourrait être FOXA1. Nous avons donc testé ce postulat en tentant d'exprimer de façon ectopique cette protéine au sein des cellules MDA::ER et d'en évaluer les conséquences au niveau du cistrome ER et du transcriptome œstrogéno-dépendant. Nous avons également opéré pour le développement de ces expériences un shift technologique en analysant les cistromes des différentes protéines à l'échelle du génome par le séquençage à haut débit de nos échantillons de ChIP.

Comme présenté dans le second manuscrit en cours de finalisation qui constitue la deuxième partie des travaux effectués au cours de ma thèse, l'introduction de FOXA1 dans les MDA::ER ne permet pas de transformer les cellules MDA::ER en cellules "MCF-7-like". En effet, FOXA1 ne permet ni au cistrome ER ni au répertoire de gènes œstrogéno-dépendant des cellules MDA::ER de se rapprocher de ceux caractérisés au sein des cellules MCF-7. De plus, au contraire des cellules MCF-7, les MDA::ER::FOXA1 présentent une prolifération réduite en présence d'E2, voire même une orientation vers la mort cellulaire/apoptose. Le cistrome ER de ces cellules s'est également avéré être fortement affecté par l'expression de FOXA1, moins de sites de liaison étant identifiés, et avec une intensité de signal plus faible que dans les cellules contrôles. Arrivés à ce stade de caractérisation de nos lignées cellulaires MDA::ER::FOXA1, une étude présentant le même type de conclusions au niveau transcriptomique a été publiée (Kong et al., 2011). Les auteurs montraient de plus que l'expression conjointe de GATA3 avec FOXA1 permettait de lever ce frein prolifératif de FOXA1 sur des cellules MDA-MB231. Cependant, les cellules MDA::ER utilisées dans nos travaux exprimant GATA3, nous ne pouvions concilier nos hypothèses et résultats avec ces données. Nous avons alors cherché à déterminer quel mécanisme pouvait rendre compte de l'effet négatif de FOXA1 sur le recrutement de ER. La caractérisation des motifs enrichis au sein des sites de liaison du ER nous a permis de montrer que les ERBSs de cellules MDA::ER incluaient des motifs forkhead. En évaluant l'expression des membres de cette famille au sein de nos cellules, nous avons pu mettre en avant un profil miroir entre MCF-7/FOXA1 et MDA::ER/FOXA2. Nos travaux ont pu mettre en évidence que FOXA2 jouait un rôle important dans la réponse œstrogénique des cellules MDA::ER. Il semblerait donc que cette

protéine soit le facteur non déterminé ayant un rôle équivalent à FOXA1, caractérisé dans la première partie.

Cependant, peut-on considérer que FOXA2 est, au même titre que FOXA1, un facteur pionnier de ER ? Dans le chapitre V de l'Introduction, j'abordais déjà cette problématique de définition des facteurs pionniers. En utilisant des siRNAs dirigés contre l'expression de FOXA2, nous avons montré que cette protéine pouvait jouer un rôle en amont d'ER, mais ceci pourrait ne refléter qu'un rôle indirect, comme AP1 ou Sp1 qui sont requis pour des mobilisations indirectes du ER (Webb *et al.*, 1999; Safe, 2001; Kim *et al.*, 2003). Pareillement, même si nous avons caractérisé un lien entre la présence de FOXA2 et celle de modifications post-traductionnelles des histones associées à des enhancers actifs (H3K4me2/H3K27ac), nous ne pouvons dire si la liaison de FOXA2 en est la cause ou la conséquence. Il faudrait déterminer si i) FOXA2 possède comme FOXA1 une affinité plus importante pour de la chromatine que de l'ADN nu, une propriété à priori essentielle pour une action pionnière (Cirillo et Zareh, 1999); ii) si la réduction de l'expression de FOXA2 a réellement un effet au niveau de la structure chromatinienne au niveau de ses sites par FAIRE par exemple; iii) si cette réduction a un effet global sur l'établissement du transcriptome E2-dépendant de nos cellules MDA::ER; iii) si ER et FOXA2 sont présents en même temps sur leurs sites de liaison par Re-ChIP-seq. Notons que des expériences de Re-ChIP sur quelques sites ER/FOXA2 ont permis de valider la pertinence de cette hypothèse. Le dernier point important émergeant de cette seconde partie des résultats est la caractérisation de l'interrelation interférente existant entre la protéine FOXA1 exprimée de façon ectopique et le FOXA2 endogène des cellules MDA::ER. Nous montrons en effet que sur des sites mobilisant les trois protéines, la compétition entre FOXA1 et FOXA2 serait à l'origine d'une diminution du recrutement de ER. Cette conclusion, pour le moment basée sur des résultats obtenus sur quelques sites reste encore à confirmer sur un panel plus important de régions génomiques. De même, la présence de ces trois protéines sur cette catégorie de sites doit être confirmée par des expériences de Re-ChIP.

A l'issue du premier chapitre, nous avons également mis en évidence l'importance particulière d'une (ou plusieurs) histone méthyl-transférase(s) non caractérisée(s) (HMT X dans la figure 7 du premier manuscrit) dans l'établissement du cistrome d'ER. Il est donc intéressant de constater avec la caractérisation des effets compétiteurs FOXA1/FOXA2 que nous arrivions à des conclusions similaires: l'expression ectopique de FOXA1 diminue de manière générale les enrichissements en H3K4me2 des sites ER, et en particulier ceux

recrutant ER/FOXA1 et FOXA2 et dont la mobilisation de ER diminue. Il se pourrait donc que sur ces sites, si FOXA1 et FOXA2 présentent une affinité différentielle pour une HMT, la compétition entre ces deux protéines ait un impact sur le recrutement de cette HMT. D'un autre côté, il se pourrait que cette diminution en H3K4me2 soit au moins en partie un effet indirect de l'expression de FOXA1. En effet, nos résultats de ChIP-seq montrent que cette baisse en H3K4me2 est globale, ayant lieu sur des sites ne mobilisant pas FOXA1, et ne peut donc être associée à une compétition FOXA1/FOXA2 sur ces sites. De plus, nos analyses transcriptomiques montrent que l'expression de certaines HMTs est affectée dans les cellules MDA::ER exprimant FOXA1. Si l'on revient au modèle MCF-7, cette hypothèse est cohérente avec l'observation que la réduction de l'expression de FOXA1 dans ces cellules a clairement un effet global sur la mobilisation de ER, observée sur des sites ne mobilisant pas FOXA1 (Hurtado *et al.*, 2011). La caractérisation de cette/ces HMTs via des techniques de ChIP/Re-ChIP ainsi que d'approches protéomiques (identification par spectrométrie de masse des protéines interagissant avec FOXA2, comme réalisé dans les cellules MCF-7 avec FOXA1; J. Carroll, communication personnelle) est certainement un des points les plus intéressants vers lequel le laboratoire devra orienter son axe de recherche.

Enfin, il reste à caractériser l'importance fonctionnelle des interrelations FOXA1/FOXA2 dans l'orientation de l'état prolifératif des cellules MDA::ER: est-ce que, comme observé dans le cas de la liaison de ER sur certains sites, la réduction préalable de l'expression de FOXA2 est suffisante pour permettre à FOXA1 de réorienter le devenir cellulaire vers la prolifération ? Est-ce que la surexpression de GATA3 est malgré tout nécessaire pour passer outre la compétition entre FOXA1 et FOXA2?

BIBLIOGRAPHIE

- Ahmad K et Henikoff S (2001). Centromere are specialized replication domains in heterochromatin. *J Cell Biol* **153**:101-110.
- Aihara T, Miyoshi Y, Koyama K, Suzuki M, Takahashi E, Monden M, Nakamura Y (1998). Cloning and mapping of SMARCA5 encoding hSNF2H, a novel human homologue of Drosophila ISWI. *Cell Genet* **81**:191-193.
- Akhtar A et Gasser S.M (2007). The nuclear envelope and transcriptional contrôle. *Nat Rev Genet* **8**:507-517.
- Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, Sali A, Rout MP (2007). The molecular architecture of the nuclear pore complex. *Nature* **450**:695-701.
- Alcobia I, Dilao R et Parreira L (2000). Spatial associations of centromeres in the nuclei of hematopoietic cells: evidence for cell-type-specific organizational patterns. *Blood* **95**:1608-1615.
- Allan G.F, Ing N.H, Tsai S.Y, Tsai M-J et O'Malley B.W (1991). Synergism between steroid response and promoter elements during cell-free transcription. *J Biol Chem* **266**:5905-5910.
- Allan G.F, Leng X.H, Tsai S.Y, Weigel N.L, Edwards D.P, Tsai M-J et O'Malley B.W (1992). Hormone and anti-hormone induce distinct conformational changes wich are cental to steroid receptor activation. *J Biol Chem* **267**:19513-19520.
- Andrau J.C, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp M.G, van de Peppel J, Werner M et Holstege F.C (2006). Genome-wide location of the coactivator mediator : binding without activation and transient Cdk8 interaction on DNA. *Mol Cell* **22**:179-192.
- Aranda A et Pascual A (2001). Nuclear hormone receptors and gene expression. *Physiol Rev* **81**:1269-1304.
- Archambault J, Pan G, Dahmus G.K, Cartier M, Marshall N, Zhang S, Dahmus M.E et Greenblatt J (1998). FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxy-terminal domain of RNA polymerase IIO. *J Biol Chem* **273**:27593-27601.
- Arceci R.J, King A, Simon M.C, Orkin S.H et Wilson D.B (1993). Mouse GATA-4 a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol Cell Biol* **13**:2235-2246.
- Ashe H.L, Monks J, Wijgerde M, Fraser P et Proudfoot N.J (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**:2494-2509.
- Asselin-Labat M.L, Sutherland K.D, Barker H, Thomas R, Shackleton M, Forrest N.C, Hartley L, Robb L, Grosveld F.G, van der Wees J ... (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differenciatio. *Nat Cell Biol* **9**:201-209.
- Auble D.T, Hansen K.E, Mueller C.G.F, Lane W.S, Thorner J et Hans S (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependant mechanism. *Gene Dev* **8**:1920-1934.

- Baker S.P et Grant P.A (2007). The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* **26**:5329-5340.
- Bancaud A, Huet S, Daigle N, Mozziconacci J, Beaudouin J et Ellenberg J (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J* **28**:3785-3798.
- Baniahmad A, Kohne A.C et Renkawitz R (1992a). A transferable silencing domain is present in the thyroid hormone receptor, in the v-erbA oncogene product and in the retinoic acid receptor. *EMBO J* **11**:1015-1023.
- Baniahmad A, Tsai S.Y, O'Malley B.W et Tsai M-J (1992b). Kindred S thyroid hormone receptor is an active and constitutive silencer and a repressor for thyroid hormone and retinoic acid response. *Proc Natl Acad Sci USA* **89**:10633-10637.
- Baniahmad A, Ha I, Reinberg D, Tsai M-J et O'Malley B.W (1993). Interaction of human thyroid hormone receptor beta with transcription factor TFIIIB may mediate target gene derepression and activation by thyroid hormone. *Proc Natl Acad Sci USA* **90**:8832-8836.
- Bantignies F et Cavalli G (2011). Polycomb group proteins: repression in 3D. *Trends Genet* **27**:454-464.
- Beato M, Herrlich P et Schütz G (1995). Steroid hormone receptors: many actors in search of a plot. *Cell* **83**:851-857.
- Beck S, Sommer P, dos Santos Silva E, Blin N et Gott P (1999). Hepatocyte nuclear factor 3 (winged helix domain) activates trefoil factor gene TFF1 through a binding motif adjacent to the TATAA box. *DNA Cell Biol* **18**:157-164.
- Beckerman R et Prives C (2010). Transcriptional regulation by p53. *Cold Spring Harb Perspect Biol* **2**:a000935.
- Beisel C et Paro R (2011). Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet* **12**:123-135.
- Bell A.C et Felsenfeld G (2000). Methylation of a CTCF dependant boundary controls imprinted expression of the Igf2 gene. *Nature* **405**:482-485.
- Bentley D (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol* **11**:336-341.
- Berk A.J (1999). Activation of RNA polymerase II transcription. *Curr Opin Cell Biol* **11**:330-335.
- Bernstein B.E, Humphrey E.L, Erlich R.L, Schneider R, Bouman P, Liu J.S, Kouzarides T et Schreiber S.L (2002). Methylation of histone H3 lys 4 in coding regions of active genes. *Proc Natl Acad Sci USA* **99**:8695-8700.
- Bertolino E, Reimund B, Wildt-Perinic D et Clerc R.G (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem* **270**:31178-31188.

- Bi X (2012). Functions of chromatin remodeling factors in heterochromatin formation and maintenance. *Sci China Life Sci* **55**:89-96.
- Biggar S.R et Crabtree G.R (1999). Continuous and widespread roles for Swi-Snf complex in transcription. *EMBO J* **18**:2254-2264.
- Bird A (2000). DNA methylation patterns and epigenetics memory. *Genes Dev* **16**:6-21.
- Birney E, Stamatoyannopoulos J.A, Dutta A, Guigo R, Gingeras T.R et al., (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**:799-8116.
- Blackwood E.M et Kadonaga J.T (1998). going the distance: a current view of enhancer action. *Science* **281**:60-63.
- Blasco M.A (2005). Telomeres and human disease: Ageing, cancer and beyond. *Nat Rev Genet* **6**:611-622.
- Blobel G.A, Nakajima T, Eckner R, Montminy M et Orkin S.H (1998). CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. *Proc Natl Acad Sci USA* **95**:2061-2066.
- Blower M.D et Karpen G.H (2001). The role of drosophila CENP-A/CID in kinetochore formation, cell-cycle progression and interaction with heterochromatin. *Nat Cell Biol* **3**:730-739.
- Blower M.D, Sullivan B.A et Karpen G.H (2002). Conserved organization of centromeric chromatin in flies and human. *Dev Cell* **2**:319-330.
- Boisvert F.M, Kruhlak M.J, Box A.K, Hendzel M.J, Bazett-Jones D.P (2001). The transcription coactivator CBP is a dynamic component of the promyelocytic leukemia nuclear body. *J Cell Biol* **152**:1099-1106.
- Bonnet J, Wang Y.H, Spedale G, Atkinson R.A, Romier C, Hamiche A, Pijnappel W.W, Timmers H.T, Tora L, Devys D et Kieffer B (2010). The structural plasticity of SCA7 domains defines their differential nucleosome-binding properties. *EMBO rep* **11**:612-618.
- Bossard P et Zaret K.S (1998). GATA transcription factors as potentiators of the gut endoderm differentiation. *Dev* **125**:4909-4917.
- Bourguet W, Ruff M, Chambon P, Gronemeyer H et Moras D (1995). Crystal structure of the ligand-binding domain of the human nuclear receptor RXR- α . *Nature* **375**:377-382.
- Bowen N.J, Fujita N, Kajita M et Wade P.A (2001). Mi-2/NuRD: multiple complexes for many purposes. *Biochim Biophys Acta* **1677**:52-57.
- Boyer L.A, Latek R.R et Peterson C.L (2004). The SANT domain: a unique histone-tail-binding module? *Mol Cell Biol* **5**:1-6.
- Boyle S, Gilchrist S, Bridger J.M, Mahy N.L, Ellis J.A et Bickmore W.A (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* **10**:211-219.

- Branco M.R et Pombo A (2006). Intermingling of chromosome territories in interphase suggest role in translocations and transcription-dependant associations. *Plos Biol* **4**:e138.
- Braunschweig U, Hogan G.J, Pagie L et van Steensel B (2009). Histone H1 binding is inhibited by histones variant H3.3. *EMBO J* **28**:3635-3645.
- Brown C.J, Ballabio A, Rupert J.L, Lafreniere R.G, Grompe M, Tonlorenzi R et Willard H.F (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**:38-44.
- Bruce K, Myers F.A, Mantouvalou E, Lefevre P, Greaves I, Bonifer C, Tremethick D.J, Thorne A.W and Crane-Robinson C (2005). The replacement histone H2AZ in hyperacetylated for mis a feature of active genes in chicken. *Nucl Acides Res* **33**:5633-5639.
- Bruck N, Vitoux D, Ferry C, Duong V, Bauer A, de Thé H et Rochette-Egly C (2009). A coordinated phosphorylation cascade initiated by p38MPAK/MSK1 directs RAR α to target promoters. *EMBO J* **28**:34-47.
- Brzozowski A.M, Pike A.C, Dauter Z, Hubbard R.E, Bonn T, Engström O, Öhman L, Greene G.L, Gustafsson J-A et Carlquist M (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**:753-758.
- Buchwitz B.J, Ahmad K, Moore L.L, Roth M.B et Henikoff S (1999). A histone H3-like protein in *C.elegans*. *Nature* **401**:547-548.
- Budhram-Mahadeo V, Morris P.J, Smith M.D, Midgley C.A, Boxer L.M et Latchman D.S (1999). p53 suppresses the activation of the Bcl-2 promoter by the Brn-3a POU family transcription factor. *J Biol Chem* **274**:15237-15244.
- Bulger M et Groudine M (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* **13**:2465-2477.
- Bunone G, Briand P-A, Miksicek R.J et Picard D (1996). Activation of the unliganded estrogen receptor by EGF involves the MAP kinase pathway and direct phosphorylation. *EMBO J* **15**:2174-2183.
- Burley S.K et Roeder R.G (1996). Biochemistry and structural biology of transcription factor IID (TFIID). *Annu Rev Biochem* **65**:769-799.
- Buschbeck M, Uribealago I, Wibowo I, Rue P, Martin D, Gutierrez A, Morey L, Guigo R, Lopez-Schier H, Di Croce L (2003). The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat Struc Mol Biol* **16**:1074-1080.
- Cairns B.R, Kim Y.J, Sayre M.H, Laurent B.C et Kornberg R.D (1994). A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5 and SNF6 gene products isolated from yeast. *Proc Natl Acad Sci USA* **91**:1950-1954.
- Calvo O et Manmey J.L (2003). Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev* **17**:1321-1327.
- Cao R, Tsukada Y et Zhang Y (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**:845-854.

- Capelson M et Hetzer M.W (2009). The role of nuclear pore in gene regulation, development and disease. *EMBO rep* **10**:697-705. Erratum in: (2009) *EMBO Rep* **10**:934.
- Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer M.W (2010). Chromatin bound nuclear pore components regulate gene expression in higher eukaryote. *Cell* **140**:372-383.
- Carlson D.F, Fahrenkrug S.C, et Hackett P.P (2012). Targeting DNA with fingers and TALENs. *Molecular Therapy–Nucleic Acids* **1**,e3; doi:10.1038/mtna.2011.5.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple C.A, Taylor M.S, Engström P.G, Frith M.C, Forrest A.R, Alkema W.B, Tan S.L, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Kayatama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond S.M, Wells C.A, Orlando V, Wahlestedt C, Liu E.T, Harbers M, Kawai J, Bajic V.B, Hume D.A et Hayashizaki Y (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genet* **38**:626-635.
- Carroll J.S, Liu X.S, Brodsky A.S, Li W, Meyer C.A, Szary A.J, Eeckhoutte J, Shao W, Hestermann E.V, Geistlinger T.R, Fox E.A, Silver P.A et Brown M (2005). Chromosome wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FOXA1. *Cell* **122**:33-43.
- Carroll J.S, Meyer C.A, Song J, Li W, Geistlinger T.R, Eeckhoutte J, Brodsky A.S, Krasnickas-Keeton E, Fertuck K.C, Hall G.F, Wang Q, Bekiranov S, Sementchenko V, Fox E.A, Silver P.A, Gingeras T.R, Liu S et Brown M (2006). Genome-wide analysis of estrogen receptor binding sites. *Nature Genetics* **38**:1289-1297.
- Caterino T.L et Hayes J.J (2011). Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem Cell Biol* **89**:35-44.
- Cerf C, Lippens G, Ramakrishnan V, Muyldermans S, Segers A, Wyns L, Wodak S.J et Hallenga K (1994). Homo- and heteronuclear two dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of histone H5. *Biochem* **33**:11079-11086.
- Chadwick B.P et Willard H.F (2004). Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci USA* **101**:17450-17455.
- Chakravarti D, LaMorte V.J, Nelson M.C, Nakajima T, Schulman I.G, Juguilon H, Montminy M et Evans R.M (1996). Role of CBP/p300 in nuclear receptor signalling. *Nature* **383**:99-103.
- Chao W, Huynh K.D, Spencer R.J, Davidow L.S et Lee J.T (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. *Science* **295**:345-347.
- Charpentier A.H, Bednarek A.K, Daniel R.L, Hawkins K.A, Laflin K.J, Gaddis S, MacLeod M.C et Aldaz C.M (2000). Effects of estrogen on global gene expression: identification of novel targets of estrogen action. *Cancer Res* **60**, 5977-5983.
- Chen J.L, Attardi L.D, Verrijzer C.P, Yokomori K et Tijan R (1994). Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**:93-105.

- Chen F, Pruett-Miller S.M, Huang Y, Gjoka M, Duda K, Taunton J, Collingwood T.N, Frodin M et Davis G.D (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat methods* **8**:753-755.
- Cho D.H, Thienes C.P, Mahoney S.E, Analau E, Filippova G.N et Tapscott S.J (2005). Antisense transcription, heterochromatin et the DM1CTG repeats are constrained by CTCF. *Mol Cell* **20**:483-489.
- Cicatiello L, Scafoglio C, Altucci L, Cancemi M, Natoli G, Facchiano A, Iazzetti G, Calogero R, Biglia N, De Bortoli M *et al.* (2004). A genomic view of estrogen actions in human breast cancer cells by expression profiling of the hormone-responsive transcriptome. *J Mol Endocrinol* **32**: 719-775.
- Cirillo L.A, McPherson C.E, Bossard P, Stevens K, Cherian S, Shim E.Y, Clark K.L, Burley S.K et Zareth K.S (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucléosome. *EMBO J* **17**:244-254.
- Cirillo L.A et Zaret K.S (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* **4**:961-969.
- Cirillo L.A, Lin F.R, Cuesta I, Friedman D, Jarnik M et Zareth K.S (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FOXA) and GATA-4. *Mol Cell* **9**:279-289.
- Clark S.J, Harrison J, Paul C.L et Frommer M (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* **22**:2990-2997.
- Clausell J, Happel N, Hale T.K, Doenecke D, Beato M et Aramay R (2009). Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *Plos One* **4**:e0007243.
- Cler E, Papi G, Schultz P et Davidson I (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol Life Sci* **66**:2123-2134.
- Conaway R.C et Conaway J.W (1993). General initiation factors for RNA polymerase II. *Annu Rev Biochem* **64**:533-561.
- Conaway R.C et Conaway J.W (2012). The Mediator complex and transcription elongation. *BBA* doi: 10.1016/j.bbagr.2012.08.017.
- Conticello S.G (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**:229.
- Cooney A.J, Tsai S.Y, O'Malley B.W et Tsai M.J (1992). Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTC A response elements, allowing COUP-TF to repress hormonal induction of the vitamin D₃, thyroid hormone, and retinoic acid receptors. *Mol Cell Biol* **12**:4153-4163.
- Cooney A.J, Leng X, Tsai S.Y, O'Malley B.W et Tsai M-J (1993). Multiple mechanisms of chicken ovalbumin upstream promoter transcription factor-dependent repression of transactivation by the vitamin D, thyroid hormone, and retinoic acid receptors. *J Biol Chem* **268**:4152-4160.

- Core L.J, Waterfall J.J et Lis J.T (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**:1845-1848.
- Costanzi C et Perhson J.R (1998). Histone macro H2A is concentrated in the inactive X chromosome of female mammals. *Nature* **393**:599-601.
- Côté J, Quinn J, Workman J.L et Peterson C.L (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SNF/SWI complex. *Science* **265**:53-60.
- Couture J.F et Trievel R.C (2006). Histone modifying enzymes : encrypting an enigmatic code. *Curr Opin Struc Biol* **16**:753-760.
- Cremer M, von Hase J, Volm T, Brero A, Kreth G, Walter J, Fischer C, Solowei I, Cremer C et Cremer T (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res* **9**:541-567.
- Cremer T et Cremer C (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**:292-301.
- Creyghton M.P, Markoulaki S, Levine S.S, Hanna J, Lodate M.A, Sha K, Young R.A, Jaenisch R et Boyer L (2008). H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell* **135**:649-661.
- Creyghton M.P, Chang A.W, Welstead G.G, Kooistra T, Carey B.W, Steine E.J, Hanna J, Lodato M.A, Frampton G.M, Sharp P.A, Boyer L.A, Young R.A et Jaenisch R (2012). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**:21931-21936.
- Crijns A.P, de Graeff P, Geerts D, Ten Hoor K.A, Hollema H, van der Sluis T, Hofstra R.M, de Bock G.H, de Jong S, van der Zee A.G et de Vries E.G (2007). MEIS and PBX homeobox proteins in ovarian cancer. *Eur J Cancer* **43**:2495-2505.
- Croft J.A, Bridger J.M, Boyle S, Perry P, Teague P et Bickmore W.A (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* **145**:1119-1131.
- Daniel J.A, Torok M.S, Sun Z.W, Schieltz D, Allis C.D, Yate J.R 3rd et Grant P.A (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. *J Biol Chem* **279**:1867-1871.
- Daniel J.A, Pray-Grant M.G et Grant P.A (2005). Effector proteins for methylated histones : an expanding family. *Cell Cycle* **4**:919-926.
- Davis J.A, Takagi Y, Kornberg R.D et Asturias F.A (2002). Structure of the yeast RNA polymerase II holoenzyme: mediator conformation and polymerase interaction. *Mol Cell* **10**:409-415.
- Dekker J, Rippe K, Dekker M et Kleckner N (2002). Capturing chromosome conformation. *Science* **295**:1306-1311.
- DeLange R.J, Fambrough D.M, Smith E.L et Bonner J (1969). Calf and Pea Histone IV. II. The complete amino acid sequence of calf thymus histone IV; Presence of epsilon-N-acetyllysine. *J Biol Chem* **244**:319-34.

- De Laat W, Klous P, Kooren J, Noordermeer D, Palstra R.J, Simonis M, Splinter E et Grosveld F (2008). Three-dimensional organization of gene expression in erythroid cells. *Curr Top Dev Biol* **82**:117-139.
- De la Serna I.L et Imbalzano A.N (2002). Unfolding heterochromatin for replication. *Nat Genet* **32**:560-562.
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tussi B.K, Muller H, Ragoussis J, Wei C.L et Natoli G (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLos Biol* **8**:1-17.
- De Wit E, Greil F et van Steensel B (2007). High resolution mapping reveals links of HP1 with active and inactive chromatin components. *Plos Genet* **3**:e38.
- De Wit E et De Laat W (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev* **26**:11-24.
- DiMartino J.F, Selleri I, Traver D, Firpo M.T, Rhee J, Warnke R, O’Gorman S, Weissman I.L et Cleary M.L (2001). The HOX cofactor and proto-oncogene PBX1 is required for maintenance of definitive hematopoiesis in the fetal liver. *Blood* **98**:618-626.
- Drapkin R et Reinberg D (1994). The multifunctional TFIIF complex and transcriptional control. *Trends Biochem Sci* **19**:504-508.
- Durocher D, Charron F, Warren R, Schwartz R et Nemer M (1997). The cardiac transcription factor Nkx2-5 and GATA-4 are mutual cofactors. *EMBO J* **16**:5687-5696.
- Dvir A, Conaway R.C et Conaway J.W (1996). Prooter escape by RNA polymerase II- a role for an ATP cofactor in suppression of arrest by polymerase at promoter-proximal sites. *J Biol Chem* **271**:23352-23356.
- Dvir A, Conaway R.C et Conaway J.W (1997). A role for TFIIF in controlling the activity of early RNA polymerase II elongation complexes. *Proc Natl Acad Sci USA* **94**:9006-9010.
- Ebbert R, Birkmann A et Schüller H.J (1998). The product of the SNF2/SWI2 paralogue INO80 of *saccharomyces cerevisiae* required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Mol Microbiol* **32**:741-751.
- Eeckhoute J, Briche I, Kurowska M, Formstecher P et Laine B (2006). Hepatocyte nuclear factor 4 alpha ligand binding and F domains mediate interaction and transcriptional synergy with the pancreatic islet LIM HD transcription factor Isl 1. *J Mol Biol* **364**:567-581.
- Eeckhoute J, Krasnickas Keeton E, Lupien M, Krum S.A, Carroll J.S et Brown M (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor α expression in breast cancer. *Cancer Res* **67**:6477-6483.
- Eeckhoute J, Lupien M, Meyer C.A, Verzi M.P, Shivdasani R.A, Liu X.S et Brown M (2009). Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* **19**:372-380.
- Eirin-Lopez J.M et Ausio J (2007). H2AZ mediated genome wide chromatin specialization. *Curr Gen* **8**:59-66.

- Elfring L.K, Deuring R, McCallum C.M, Peterson C.L et Tamkun J.W (1994). Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Mol Cell Biol* **14**:2225-2234.
- Elliston J.F, Fawell S.E, Klein-Hitpass L, Tsai S.Y, Tsai M-J, Parker M.G et O'Malley B.W (1990). Mechanism of estrogen receptor-dependant transcription in a cell-free system. *Mol Cell Biol* **10**:6607-6612.
- Ernst J et Kellis M (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology* **28**:817-825.
- Faast R, Thonglairoam V, Schulz T.C, Beall J, Wells J.R, Taylor H, Matthaei K, Rathjen P.D, Tremethick D.J et Lyons I (2001). Histone variant H2AZ is required for early mammalian development. *Curr Biol* **11**:1183-1187.
- Felsenfeld G (1992). Chromatin as an essential part of the transcriptional mechanisms. *Nature* **355**:219-225.
- Filion G.J, van Bommel J.G, Braunschweig U, Talhout W, Kind J, Ward L.D, Brugman W, de Castro I.J, Kerkhoven R.M, Bussemaker H.J et van Steensel B (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**:212-224.
- Filippova G.N, Fagerlie S, Klenova E.M, Myers C, Dehner Y, Goodwin G, Neiman P.E, Collins S.J et Lobanenko V.V (1996). An exceptionally conserved transcriptional repressor CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* **16**:2802-2813.
- Filippova G.N (2008). Genetics and epigenetics of the multi-functional protein CTCF. *Curr Top Dev Biol* **80**:337-360.
- Finlan L.E, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR, Bickmore W.A (2008). Recruitment to the nuclear periphery can alter expression of gene in human cells. *PLoS Genet* **4**:e1000039.
- Fish R.N et Kane C.M (2002). Promoting elongation with transcript cleavage stimulatory factors. *BBA* **1577**:287-307.
- Flouriou G, Brand H, Denger S, Metivier R, Kos M, Reid G, Soontag-Buck V et Gannon F (2000). Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1. *EMBO J* **19**:4688-4700.
- Fondell J.D, Roy A.L et Roeder R.G (1993). Unliganded thyroid hormone receptor inhibits formation of a functional preinitiation complex: implications for active repression. *Genes Dev* **7**:1400-1410.
- Fritz E.L et Papavasiliou N (2010). Cytidine deaminases : AIDing DNA methylation ? *Genes Dev* **24**:2107-2114.
- Fu Y, Sinha M, Peterson C.L et Weng Z (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* **25**:e1000138.

- Fullwood M.J, Wei C.L, Liu E.T et Ruan Y (2009a). Next generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res* **19**:521-532.
- Fullwood M.J, Liu M.H, Pan Y.F, Liu J, Xu H, Mohamed YB, Orlov Y.L, Velkov S, Ho A, Mei P.H, Chew E.G, Huang P.Y, Welboren W.J, Han Y, Ooi H.S, Ariyaratne P.N, Vega V.B, Luo Y, Tan P.Y, Choy P.Y, Wansa K.D, Zhao B, Lim KS, Leow S.C, Yow J.S, Joseph R, Li H, Desai K.V, Thomsen J.S, Lee Y.K, Karuturi R.K, Herve T, Bourque G, Stunnenberg H.G, Ruan X, Cacheux-Rataboul V, Sung W.K, Liu E.T, Wei C.L, Cheung E et Ruan Y (2009b). An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**:58-64.
- Gamble M.J, Frizzell K.M, Yang C, Krishnakumar R et Kraus L (2010). The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes Dev* **24**:21-32.
- Gao X, Sedgwick T, Shi Y.B et Evans T (1998). Distinct functions are implicated for the GATA-4, -5, and -6 transcription factors in the regulation of intestine epithelial cell differentiation. *Mol Cell Biol* **18**:2901-2911.
- Gao N, Zhang J, Rao M.A, Case T.C, Mirosevich J, Wang Y, Jin R, Gupta A, Rennie P.S et Matusik R.J (2003). The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol Endocrinol* **17**:1484-1507.
- Gao N, Lelay J, Vatamaniuk M.Z, Rieck S, Friedman J.R et Kaestner K.H (2008). Dynamic regulation of PDX1 enhancers by FoxA1 and FOXA2 is essential for pancreas development. *Genes Dev* **22**:3435-3448.
- Garcia M, Derocq D, Freiss G et Rochefort H (1992). Activation of estrogen receptor transfected into a receptor-negative breast cancer cell line decreases the metastatic and invasive potential of the cells. *Proc Natl Acad Sci USA* **89**:11538-11542.
- Gaszner M et Felsenfeld G (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7**:703-713.
- Gerasimova T.I, Byrd K et Corces V.G (2000). A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* **6**:1025-1035.
- Gershenson N.I et Ioshikhes I.P (2005). Synergy of human pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* **21**:1295-1300.
- Gerstein M.B, Lu Z.J, Van ostrand E.L, Cheng C, Arshinoff B.I, Liu T, Yip K.Y, Robilotto R, Rechtsteiner A, Ikegami K, Alves P et al., (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**:1775-1787.
- Geyer P.K, Green M.M et Corces V.G (1990). Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in *Drosophila*. *EMBO J* **9**:2247-2256.
- Gill G, Pascal E, Tseng Z.H et Tjian R (1994). A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *drosophila* TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci USA* **91**:192-196.

- Goldberg A.D, Banaszynski L.A, Noh K.M, Lewiw P.W, Elsaesser S.J, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chapgier A, DeKelver R.C, Miller J.C, Lee Y.L, Boydston E.A, Holmes M.C, Gregory P.D, Greally J.M, Rafii S, Yang C, Scambler P.J, Garrick D, Gibbons R.J, Higgs D.R, Cristea I.M, Urnov F.D, Zheng D et Allis C.D (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**:678-691.
- Gorski S.A, Dundr M et Misteli T (2006). The road much traveled : trafficking in the cell nucleus. *Curr Opin Cell Biol* **18**:284-290.
- Gorski S.A, Snyder S.K, John S, Grummt I et Misteli T (2008). Modulation of RNA polymerase assembly dynamics in transcriptional regulation. *Mol Cell* **30**:486-497.
- Govind C.K, Zhang F, Qiu H, Hofmeyer K et Hinnebusch A.G (2007). Gcn5 promotes acetylation, eviction, and méthylation of nucleosomes in transcribed coding regions. *Mol Cell* **25**:31-42.
- Gowher H, Leismann O et Jeltsch A (2000). DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J* **19**:6918-6923.
- Grande M.A, van der Kraan I, de Jong L, van Driel R (1997). Nuclear distribution of transcription factors relation to sites of transcription and RNA polymerase II. *J Cell Sci* **110**:1781-1791.
- Gribnau J, Diderich K, Pruzina S, Calzolari R et Fraser P (2000). Intergenic transcription and developemtal remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* **5**:377-386.
- Gronemeyer I.I et Laudet V (1995). Transcription factor 3: nuclear receptor. *Protein Profile* **2**:1173-1308.
- Gruenbaum Y, Goldman R.D, Meyuhas R, Mills E, Margalit A, Fridkin A, Dayani Y, Prokocimer M et Enosh A (2003). The nuclear lamina and its functions in the nucleus. *Int Rev Cytol* **226**:1-62.
- Grüne T, Brzeski J, Eberharter A, Clapier C.R, Corona D.F, Becker P.B, Müller C.W (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell* **12**:449-460.
- Guillemette B et Gaudreau L (2006). Reuniting the contrasting functions of H2A.Z. *Biochem Cell Biol* **84**:528-535.
- Gullerova M et Proudfoot N.J (2008). Cohesin complex promotes trabscriptionnal termination between convergent genes in *S.pombe*. *Cell* **132**:983-995.
- Guo J.U, Su Y, Zhong C, Ming G.L et Song H (2011). Hydrosylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**:423-434.
- Hadjur S, Williams L.M, Ryan N.K, Cobb B.S, Sexton T, Fraser P, Fisher A.G et Merckenschlager M (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**:410-414.
- Hager G.L, McNally J.G et Misteli T (2009). Transcription Dynamics. *Mol Cell* **35**:741-753.
- Hah N, Danko C.G, Core L, Waterfall J.J, Siepel A, Lis J.T et Kraus L.W (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145**:622-634.

- Hahn S (1998). The role of TAFs in RNA polymerase II transcription. *Cell* **95**:579-582.
- Hahn S (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**:394-403.
- Hajkova P, Erhardt S, Lane S, Haaf T, El-Maarri O, Reik W, Walter J et Surani M.A (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**:15-23.
- Hake S.B et Allis C.D (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the “H3 barcode hypothesis”. *Proc Natl Acad Sci USA* **103**:6428-6435.
- Handoko L, Xu H, Li G, Ngan C.Y, Chew E, Schnapp M, Lee C.W, Ye C, Ping J.L, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan C.S, Kunarso G, Shahab A, Bourque G, Cacheux-Rataboul V, Sung W.K, Ruan Y et Wei C.L (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genet* **43**:630-638.
- Hanstein B, Eckner R, DiRenzo J, Halachmi S, Liu H, Searcy B, Kurokawa R et Brown M (1996). p300 is a component of an estrogen receptor coactivator complex. *Proc Natl Acad Sci USA* **93**:11540-11545.
- Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997). Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* **15**:345-355.
- Hartzog G.A et Fu J (2012). The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *BBA* doi: 10.1016/j.bbagr.2012.08.007.
- Heard E et Disteche C.M (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**:1848-1867.
- Heintzman N.D, Stuart R.K, Hon G, Fu Y, Ching C.W, Hawkins R.D, Barrera L.O, Van Calcar S, Qu C, Ching K.A, Wang W, Weng Z, Green R.D, Crawford G.E et Ren B (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**:311-318.
- Heintzman N.D, Stuart R.K, Hon G, Fu Y, Ching C.W, Hawkins R.D, Barrera L.O, Van Calcar S, Qu C, Ching K.A, Wang W, Weng Z, Green R.D, Crawford G.E et Ren B (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**:311-318.
- Heintzman N.D, Hon G.C, Hawkins R.D, Kheradpour P, Stark A, Harp L.F, Ye Z, Lee L.K, Stuart R.K, Ching C.W, Ching K.A, Antosiewicz-Bourget J.E, Liu H, Zhang X, Green R.D, Lobanenkov V.V, Stewart R, Thomson J.A, Crawford G.E, Kellis M et Ren B (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**:108-112.
- Henikoff S, Ahmad K, Platero J.S et van Steensel B (2000). heterochromatic deposition of centromeric histone H3-like proteins. *Proc Natl Acad Sci USA* **97**:716-721.
- Hernandez N (1993). TBP, a universal eukaryotic transcription factor? *Genes Dev* **7**:1291-1308.
- Hirano T (2005). SMC proteins and chromosome mechanics: From bacteria to humans. *Philos Trans R Soc Lond B Biol Sci* **360**:507-514.

- Ho J.S, Ma W, Mao D.Y et Benchimol S (2005). p53-dependant transcriptional repression of c-myc is required for G1 cell cycle arrest. *Mol Cell Biol* **25**:7423-7431.
- Ho L et Crabtree G.R (2010). Chromatin remodeling during development. *Nature* **463**:474-484.
- Hoffman W.H, Biade S, Zilfou J.T, Chen J et Murphy M (2002). Transcriptional repression of the anti-apoptotic survivin gene by wild type p53. *J Biol Chem* **277**:3247-3257.
- Hoffman B.G, Robertson G, Zavaglia B, Beach M, Cullum R, Lee S, Soukhatcheva G, Li L, Wederell E.D, Thiessen N, Bilenki M, Cezard T, Tam A, Kamoh B, Birol I, Dai D, Zhao Y, Hirst M, Verchere C.B, Helgason C.D, Marra M.A, Jones S.J et Hoodless P.A (2010). Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PBX1-bound loci in islets and liver. *Genome Res* **20**:1037-1051.
- Holaska J.M, Lee K.K, Kowalski A.K et Wilson K.L (2003). Transcriptional repressor germ cell-less (GCL) and barrier to autointegration factor (BAF) compete for binding to emerlin in vitro. *J Biol Chem* **278**:6969-6975.
- Holmqvist P-H, Belikov S, Zaret K.S et Wrangé Ö (2005). FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. *Exp Cell Res* **304**:593-603.
- Hong H, Kohli K, Trivedi A, Johnson D.L et Stallcup M.R (1996). Grip1, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone-binding domains of steroid-receptors. *Proc Natl Acad Sci USA* **93**:4948-4952.
- Horlein A.J, Naar A.M, Heinzel T, Torchia J, Gloss B, Kurokawa R, Ryan A, Kamel Y, Soderstrom M, Glass C.K et Rosenfeld M.G (1995). Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear co-repressor. *Nature* **377**:397-404.
- Horn P.J, Carruthers L.M, Logie C, Hill D.A Solomon M.J, Wade P.A, Imbalzano A.N Hansen J.C et Peterson C.L (2002). Phosphorylation of linker histones regulates ATP*dependant chromatin remodeling enzymes. *Nat Struct Biol* **9**:263-267.
- Hou C, Dale R et Dean A (2009). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *PNAS* **107**: 3651-3656.
- Howman E.V, Fowler K.J, Newson A.J, Redward S, MacDonald A.C, Kalitsis P et Choo K.H (2000). Early disruption of centromeric chromatin organisation in centromere protein A null mice. *Proc Natl Acad Sci USA* **97**:1148-1153.
- Hu J.M, Camper S.A, Tilghman S.M, Miller T, Gerogoff I, Serra R et Isom H.C (1992). Functional analyses of albumin expression in a series of hepatocyte cell lines and in primary hepatocytes. *Cell Growth Differ* **3**:577-588.
- Hu Q, Kwon Y.S, Nunez E, Cardamone M.D, Hutt K.R, Ohgi K.A, Garcia-Bassets I, Rose D.W, Glass C.K, Rosenfeld M.G et Fu X.D (2008). Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *PNAS* **105**:19199-19204.
- Huang S et Spector D.L (1996). Dynamic organization of pre-mRNA splicing factors. *J Cell Biochem* **62**:191-197.

- Humphrey G.W, Wang Y, Russanova V.R, Hirai T, Qin J, Nakatani Y et Howard B.H (2001). Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *Biol Chem* **276**:6817-6824.
- Hurtado A, Holmes K.A, Ross-Innes C.S, Schmidt D et Carroll J.S (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43**:27-33.
- Huynh V.A, Robinson P.J et Rhodes (2005). A methode for the in vitro reconstitution of a defined « 30nm » chromatin fibers containing stoichiometric amounts of the linker histone. *J Mol Biol* **345**:957-968.
- Imbalzano A.N, Kwon H, Green M.R et Kingston R.E (1994). Facilitated binding of TATA-binding protein to nucleosomal DNA. *Nature* **370**:481-485.
- Ing N.H, Beekman J.M, Tsai S.Y, Tsai M-J et O'Malley B.W (1992). Members of the steroid hormone receptor superfamily interact with TFIIB (S300-II). *J Biol Chem* **267**:17617-17623.
- Ishii K, Arib G, Lin C, Van Houwe G et Laemmli U.K (2002). Chromatin boundaries in budding yeast: the nuclear pore connection. *Cell* **109**:551-562.
- Ito T, Bulger M, Pazin M.J, Kobayashi R et Kadonage J.T (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**:145-155.
- Ito S, Shen L, Dai Q, Wu S.C, Collins L.B, Swenberg J.A, He C et Zhang Y (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**:1299-1203.
- Jackson D.A, Rowader K.E, Stevens K, Jiang C, Milos P et Zaret K.S (1993). Modulation of liver-specific transcription by interactions between hepatocyte nuclear factor 3 and nuclear factor 1 binding DNA in close apposition. *Mol Cell Biol* **13**:2401-2410.
- Janowski B.A, Willy P.J, Rama Devi T, Falck J.R et Mangelsdorf D.J (1996). An oxysterol signalling pathway mediated by the nuclear receptor LXR α . *Nature* **383**:728-731.
- Jimenez-Garcia J.F et Spector D.L (1993). In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell* **73**:47-59.
- Jin C et Felsenfeld G (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* **21**:1519-1529.
- Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K et Felsenfeld G (2009). H3.3/H2A.Z double variant-containing nucleosomes mark «nucleosome-free regions» of active promoters and other regulatory regions. *Nat Genet* **41**:941-945.
- Jing H, Vakoc C.R, Ying L, Mandat S, Wang H, Zheng X et Blobel G.A (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Mol Cell* **29**:232-242.
- Jones P.A et Takai D (2001). The role of DNA methylation in Mammalian epigenetics. *Science* **293**:1067-1070.

- Kamei Y, Xu L, Heinzel T, Torchia J, Kurokawa R, Gloss B, Lin S.C, Heyman R.A, Rose D.W, Glass C.K et Rosenfeld M.G (1996). CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* **85**:403-414.
- Kang Z, Pirskanen A, Janne O.A et Palvimo J.J (2002). Involvement of proteasome in the dynamic assembly of the androgen receptor transcription complex. *J Biol Chem* **277**:48336-48371.
- Karin M (1998). New twists in gene regulation by glucocorticoid receptor: is DNA binding dispensable ? *Cell* **93**:487-490.
- Kato S, Endoh H, Masuhiro Y, Kitamoto T, Uchiyama S, Sasaki H, Masushige S, Gotoh Y, Nishida E, Kawashima H, Metzger D et Chambon P (1995). Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270**:1491-1494.
- Katoh M (2004). Human FOX gene family. *Int J Oncol* **25**:1495-1500.
- Katznellenbogen J.A et Katznellenbogen B.S (1996). Nuclear hormone receptors: ligand-activated regulators of transcription and diverse cell responses. *Chem Biol* **3**:529-536.
- Katznellenbogen B.S et Korach K.S (1997). Editorial: a new actor in the estrogen receptor drama-enter ER β . *Endocrinol* **138**:861-862.
- Katznellenbogen B.S, Choi I, Delage-Mouroux R, Ediger T.R, Martini P.G, Montano M, Sun J, Weis K et Katznellenbogen J.A (2000). Molecular mechanisms of estrogen action: selective ligands and receptor pharmacology. *J Ste Biochem & Mol Biol* **74**:279-285.
- Kelly M.J et Levin E.R (2001). Rapid actions of plasma membrane estrogen receptors. *TRENDS endocrinol et metabolism* **12**:152-156.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**:480-485.
- Khavari P.A, Peterson C.L, Tamkun J.W et Crabtree G.R (1993). BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**:170-174.
- Kikugawa T, Kinugasa Y, Shiraishi K, Nanba D, Nakashiro K, Tanji N, Yokoyama M et Higashivama S (2006). PLZF regulated Pbx1 transcription and Pbx1-HoxC8 complex leads to androgen-independent prostate cancer proliferation. *Prostate* **66**:1092-1099.
- Kim Y.J, Bjorklund S, Li Y, Sayre M et Kornberg R (1994). A multiportein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**:599-608.
- Kim T.H, Abdullaev Z.K, Smith A.D, Ching K.A, Loukinov D.I, Green R.D, Zhang M.Q, Lobanenko V.V et Ren B (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**:1231-1245.
- Kim S.I, Bultman S.J, Kiefer C.M, Dean A et Bresnick E.H (2009). BRG1 requirement for long-range interaction of a locus contrl region with a downstream promoter. *Proc Natl Acad Sci USA* **106**:2259-2264.

- Kim T.K, Hemberg M, Gray J.M, Costa A.M, Bear D.M, Wu J, Harmin D.A, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley P.F, Kreiman G et Greenberg M.E (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**:182-187.
- Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi K, Isogai T et Sugano S (2006). diversification of transcriptional modulation : large scale identification and characterization of putative alternative promoters. *Genome Res* **16**:55-65.
- Kingston R, Bunker C et Imbalzano A.N (1996). Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev* **10**:905-920.
- Klein-Hitpass L, Tsai S.Y, Weigel N.L, Allan G.F, Riley D, Rodriguez R, Schrader W.T, Tsai M-J et O'Malley B.W (1990). The progesterone receptor stimulates cell-free transcription by enhancing the formation of a stable preinitiation complex. *Cell* **60**:247-257.
- Kliwer S.A, Umesono K, Noonan D.J, Heyman R.A et Evans R.M (1992). Convergence of 9-cis retinoic acid and peroxisome proliferator signalling pathway through heterodimer formation of their receptors. *Nature* **358**:771-774.
- Kocanova S, Kerr E.A, Rafique S, Boyle S, Katz E, Caze-Subra S, Bickmore W.A et Bystricky K (2010). Activation of estrogen-responsive genes does not require their nuclear colocalization. *PLoS Genetics* **6**:e1000922.
- Koch C.M, Andrews R.M, Flicek P, Dillon S.C, Karaös U, Clelland G.K, Wilcox S, Beare D.M, Fowler J.C, Couttet P, James K.D, Lefebvre G.C, Bruce A.W, Dovey O.M, Ellis P.D, Dhami P, Langford C.F, Weng Z, Birney E, Carter N.P, Vetrie D et Dunham I (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* **17**:691-707.
- Kohler S et Cirillo L.A (2010). Stable chromatin binding prevents FoxA acetylation, preserving FoxA chromatin remodeling. *J Biol Chem* **285**:464-472.
- Koleske A.J et Young R.A (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**:466-469.
- Kong S.L, Li G, Loh S.L, Sung W-K et Liu E.T (2011). Cellular reprogramming by the conjoint action of ER α , FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol* **7**:526.
- Kornberg R.D (2005). Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* **30**:235-239.
- Kouros-Mehr H, Slorach E.M, Sternlicht M.D et Werb Z (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**:1041-1055.
- Koutelou E, Hirsch C.L et Dent S.Y (2010). Multiple faces of the SAGA complex. *Curr Opin Cell Biol* **22**:374-382.

- Kriaucionis S et Heintz N (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**:928-930.
- Krum S.A, Miranda-Carboni G.A, Lupien M, Eeckhoutte J, Carroll J.S et Brown M (2008). Unique ERalpha cisomes control cell type-specific gene regulation. *Mol Endocrinol* **22**:2393-2406.
- Ku M, Koche R.P, Rheinbay E, Mendenhall E.M, Endoh M, Mikkelsen T.S, Presser A, Nusbaum C, Xie X, Chi A.S, Adli M, Kasif S, Ptaszek L.M, Cowan C.A, Lander E.S, Koseki H, Nernstein B.E (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**:e1000242.
- Kuiper G.J, Enmark E, Pelt-Huillo M, Nilsson S et Gustafsson J.A (1996). Cloning of a novel estrogen receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci USA* **93**:5925-5930.
- Kuiper G.J, Carlquist M et Gustafsson J.A (1998). Estrogen is a male and female hormone. *Sci Med* **5**:36-45.
- Kusch T et Workman J.L (2007). Histone variants and complexes involved in their exchange. *Subcell Biochem* **41**:91-109.
- Kustatscher G, Hothorn M, Pugieux C, Scheffzek K et Ladurner A.G (2005). Splicing regulates NAD metabolite binding to histone MacroH2A. *Nat Struct Mol Biol* **12**:624-625.
- Kwon H, Imbalzano A.N, Khavari P.A, Kingston R.E et Green M.R (1994). Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* **370**:477-481.
- Lacroix M et Leclercq G (2004). About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. *Mol Cell Endocrinol* **219**:1-7.
- Ladurner A.G (2003). Inactivating chromosomes: A macro domain that minimizes transcription. *Mol Cell* **12**:1-3.
- Laganiere J, Deblois G, Lefebvre G, Bataille A.R, Robert F et Giguere V (2005). From the cover: location analysis of estrogen receptor alpha target promoters reveals that FOA1 defines a domain of the estrogen response. *Proc Natl Acad Sci USA* **102**:11651-11656.
- Lai E, Prezioso V.R, Smith E, Litvin O, Costa R.H et Darnell J.E Jr (1990). HNF-3A, a hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes Dev* **4**:1427-1436.
- Lai E, Prezioso V.R, Tao W, Chen W.S et Darnell J.E Jr (1991). Hepatocyte nuclear factor 3A belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. *Genes Dev* **5**:416-427.
- LaMorte V.J, Dyck J.A, Ochs R.L, Evans R.M (1998). Localization of nascent RNA and CREB binding protein with the PML-containing nuclear body. *Proc Natl Acad Sci* **95**:4991-4996.
- Lan X, Witt H, Katsumura K, Ye Z, Wang Q, Bresnick E.H, Farnham P.J et Jin V.X (2009). Integration of Hi-C and CHIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res* **40**:1-15.

- Lange M, Demajo S, Jain P et Di Croce L (2011). Combinatorial assembly and function of chromatin regulatory complexes. *Epigenomics* **3**:567-580.
- Laudet V (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J Mol Endocrinol* **19**:207-226.
- Laurent B.C, Treich I et Carlson M (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev* **7**:583-591.
- Laverriere A.C, MacNeill C, Mueller C, Poelmann R, Burch J et Evans T (1994). GATA-4/5/6 a subfamily of three transcription factors transcribed in developing heart and gut. *J Biol Chem* **269**:23177-23184.
- Lee M.S, Kliewer S.A, Provencal J, Wright P.E et Evans R.M (1993). Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding. *Science* **260**:1117-1121.
- Lee J.T, Davidow L.S et Warshawsky D (1999). Tsix, a gene antisens to Xist at the X-inactivation centre. *Nat Genet* **21**:400-404.
- Lee T.I et Young R.A (2000). Transcription of eukaryotic protein coding genes. *Annu Rev Genet* **34**:77-137.
- Lee C.S, Friedman J.R, Fulmer J.T et Kaestner K.H (2005). The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**:944-947.
- Lee K.K, Sardi M.E, Swanson S.K, Gilmore J.M, Torok M, Grant P.A, Florens L, Workman J.L et Washburn M.P (2011). Combinatorial depletion analysis to assemble the network architecture of the SAGA and ADA chromatin remodeling complexes. *Mol Syst Biol* **7**:503.
- Leid M, Kastner P et Chambon P (1992). Multiplicity generates diversity in the retinoic acid signalling pathways. *Trends in biochemical science* **17**:427-433.
- Lemaire V, Lee C.F, Metivier R et Glass L (2006). Sequential recruitment and combinatorial assembling of multiprotein complexes in transcriptional activation. *Phys Rev Lett* **96**:198102
- Leonard M.W, Lim K.C et Engel J.D (1993). Expression of the chicken GATA factor family during early erythroid development and differentiation. *Dev* **119**:519-531.
- Li C, Briggs M.R, Ahlborn T.E, Kraemer F.B et Liu J (2001). Requirement of Sp1 and estrogen receptor alpha interaction in the 17beta-estradiol-mediated transcriptional activation of the low density lipoprotein receptor gene expression. *Endocrinology* **142**:1546-1553.
- Li Y, Sun L, Zhang Y, Wang D, Wang F, Liang J, Gui B et Shang Y (2010). The histone modifications governing TFF1 transcription mediated by estrogen receptor. *J Biol Chem* **286**:13925-13936.
- Lieberman-Aiden E, van Berkum N.L, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie B.R, Sabo P.J, Dorschner M.O, Sandstrom R, Bernstein B, Bender M.A, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny L.A, Lander E.S et Dekker J (2009). Comprehensive mapping of long range interactions reveals folding principles of the Human genome. *Science* **326**:289-293.

- Lin Y.C, Choi W.S et Gralla J.D (2005). TFIID XBP mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat Struct Mol Biol* **12**:603-607.
- Lin C.Y, Vega V.B, Thomsen J.S, Zhang T, Kong S.L, Xie M, Chiu K.P, Lipovitch L, Barnett D.H, Stossi F, Yeo A, George J, Kuznetsov V.A, Lee Y.K, Charn T.H, Palanisamy N, Miller L.D, Cheung E, Katzenellenbogen B.S, Ruan Y, Bourque G, Wei C.L et Liu E.T (2007). Whole genome cartography of estrogen receptor alpha binding sites. *PLoS One* **3**:e87.
- Lister R, Pelizzola M, Dowen R.H, Hawkins R.D, Hon G, Tonti-Filippini J, Nery J.R, Lee L, Ye Z, Ngo Q.M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar A.H, Thomson J.A, Ren B et Ecker J.R (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**:315-322.
- Liu J-K, DiPersio C.M et Zaret K.S (1991). Extracellular signals that regulate liver transcription factors during hepatic differentiation in vitro. *Mol Cell Biol* **11**:773-784.
- Liu Y, Ranish J.A, Aebersold R et Hahn S (2001). Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. *J Biol Chem* **276**:7169-7175.
- Liu D.B, Gu Z.D, Cao X.Z, Liu H et Li J.Y (2005). Immunocytochemical detection of HoxD9 and Pbx1 homeodomain protein expression in Chinese esophageal squamous cell carcinomas. *World J Gastroenterol* **11**:1562-1566.
- Liu Y, Xia X, Fondell J.D et Yen P.M (2006). Thyroid hormone-regulated target genes have distinct patterns of coactivator recruitment and histone acetylation. *Mol Endocrinol* **20**:483-490.
- Liu J et Krantz I.D (2009). Cornelia de Lange syndrome, cohesin and beyond. *Clin Genet* **76**:303-314.
- Lo A.W, Craig J.M, Saferry R, Kalitsis P, Irvine D.V, Earle E, Magliano D.J et Choo K.H (2001). A 330 KB CENP-A binding domain and altered replication timing at human neocentromere. *EMBO J* **20**:2087-2096.
- Lobanenkov V.V, Nicolas R.H, Adler V.V, Paterson H, Klenova E.M, Polotskaja A.V et Goodwin G.H (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequences of the chicken c-myc gene. *Oncogene* **5**:743-753.
- Lopez-Soler R.I, Moir R.D, Spann T.P, Stick R, Goldman R.D (2001). A role for nuclear lamin in nuclear envelope assembly. *J Cell Biol* **154**:61-70.
- Luger K et Hansen J (2005). Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol* **15**:188-196.
- Lupien M, Eeckhoute J, Meyer C.A, Wang Q, Zhang Y, Li W, Carroll J.S, Liu X.S et Brown M (2008). FOXA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**:958-970.
- Lyko F, Ramsahoye B.H et Jaenisch R (2000). DNA methylation in *Drosophila melanogaster*. *Nature* **408**:538-540.
- McNally J.G, Mueller W.G, Walker D, Wolford R.G et Hager G.L (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* **287**:1262-1265.

- McPherson C.E, Horowitz R, Woodcock C.L, Jiang C et Zaret K.S (1996). Nucleosome positioning properties of the albumin transcriptional enhancer. *Nucleic Acids Res* **24**:397-404.
- Magnani L, Ballantyne E.B, Zhang X et Lupien M (2011a). PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS genet* **7**:e1002368.
- Mahy N.L, Perry P.E et Bickmore W.A (2002a). Gene density and transcription influence the localization of chromatin outside of chromosomes territories detectable by FISH. *J Cell Biol* **159**:753-763.
- Mahy N.L, Perry P.E, Gilchrist S, Baldock R.A et Bickmore W.A (2002b). Spatial organization of active and inactive genes and non coding DNA within chromosome territories. *J Cell Biol* **157**:579-589.
- Maiyar A.C, Phu P.T, Huang A.J et Firestone G.L (1997). Repression of glucocorticoid receptor transactivation and DNA binding of a glucocorticoid response element within the serum/glucocorticoid-inducible protein kinase (sgk) gene promoter by the p53 tumor suppressor tumor. *Mol Endocrinol* **11**:312-329.
- Maldonado E et Reinberg D (1995). News on initiation and elongation of transcription by RNA polymerase II. *Curr Opin Cell Biol* **7**:352-361.
- Malik H.S et Henikoff S (2003). Phylogenomics of the nucleosome. *Nat Struc Biol* **10**:882-891.
- Mangelsdorf D.J, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P et Evans R.M (1995). The nuclear receptor superfamily: the second decade. *Cell* **83**:835-839.
- Margueron R, Trojer P et Reinberg D (2005). The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**:163-176.
- Martin C et Zhang Y (2005). The diverse functions of histone lysine méthylation. *Nat Rev Mol Cell Biol* **16**:753-760.
- Massie C.E et Mills I.G (2008). ChIPping away at gene regulation. *EMBO rep* **9**:337-343.
- Matera A.G, Frey M.R, Margelot K, Wolin S.L (1995). A perinuclear compartment contains several RNA pol III transcripts as well as the polypyrimidine tract binding protein, hnRNPI. *J Cell Biol*. **129**:1181-1193.
- Mayer W, Niveleau A, Walter J, Fundele R et Haaf T (2000). Demethylation of the zygotic paternal genome. *Nature* **403**:501-502.
- Meinhart A, Kamenski T, Hoepfner S, Baumli S et Cramer P (2005). A structural perspective of CTD function. *Genes Dev* **19**:1401-1415.
- Menasce L.P, White G.R, Harrison C.J et Boyle J.M (1993). Localization of the estrogen receptor locus (ESR) to chromosome 6q25.1 by FISH and a simple post-FISH banding technique. *Genomics* **17**:263-265.
- Menendez D, Inga A et Resnick M.A (2009). The expanding universe of p53 targets. *Nat Rev Cancer* **9**:724-737.

- Merika M, Williams A.J, Chen G, Collins T et Thanos D (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergetic activation of transcription. *Mol Cell* **1**:277-287.
- Metivier R, Penot G, Hübner M.R, Reid G, Brand H, Koš M et Gannon F (2003). Estrigen receptor- α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**:751-763.
- Michaelis C, Ciosk R et Nasmyth K (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatid. *Cell* **91**:35-45.
- Migliaccion A, Di Domenico M, Castoria G, de Falco A, Bontempo P, Nola E et Auricchio F (1996). Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. *EMBO J* **15**:1292-1300.
- Millar C.B, Guy J, Sansom O.J, Selfridge J, MacDougall E, Hendrich B, Keightley P.D, Bishop S.M, Clarke A.R et Bird A (2002). Enhancer CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**:403-405.
- Misteli T, Gunjan A, Hock R, Bustin M et Brown D.T (2000). Dynamic binding of histone H1 to chromatin in living cells. *Nature* **408**:877-881.
- Misteli T (2005). Concept in nuclear architecture. *Bioessays* **27**:477-487.
- Mito Y, Henikoff J.G et Henikoff S (2007). Histone replacement marks the boudaries of cis-regulatory domains. *Science* **315**:1408-1411.
- Mizuguchi G, Shen X, Landry J, Wu W.H, Sen S et Wu C (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**:343-348.
- Mohn F, Weber M, Rebhan M, Roloff T.C, Richter J, Stadler M.B, Bibel M et Schübeler D (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**:755-766.
- Moreland R.J, Tirode F, Yan Q, Conaway W, Egly J.M et Conaway R.C (1999). A role for the TFIHH XPB DNA helicase in promoter escape by RNA polymerase II. *J Biol Chem* **274**:22127-22130.
- Morgan H.D, Dean W, Coker H.A, Reik W et Petersen-Mahrt S.K (2004). Activation-induced cytidine demaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* **279**:52353-52360.
- Morillon A, Karabetsou N, O'Sullivan J, Kent N, Proudfoot N et Mellor J (2003). ISWI chromatin remodeling ATPase corrdinates transcription elongation and termination by RNA polymerase II. *Cell* **115**:425-435.
- Mosselman S, Polman J et Dijkema R (1996). ER β : identification and characterization of a novel human estrogen receptor. *FEBS Lett* **392**:49-53.
- Muchardt C et Yanic M (1993). A human homologue of saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoide receptor. *EMBO J* **12**:4279-4290.

- Nativio R, Wendt K.S, Ito Y, Huddleston J.E, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters J.M et Murrell A (2008). Cohesin is required for higher order chromatin conformation at the imprinted IGF2/H19 locus. *Plos Genet* **5**:e1000739.
- Nguyen T.T, Cho K, Stratton S.A et Barton M.C (2005). Transcription factor interactions and chromatin modifications associated with p53-mediated, developmental repression of the alpha-fetoprotein gene. *Mol Cell Biol* **25**:2147-2157.
- Nikolov D et Burley S.K (1991). 2-1 A resolution refined structure of a TATA box-binding protein (TBP). *Nat Struc Biol* **1**:621-937.
- Nili E, Cojocaru G.S, Kalma Y, Ginsberg D, Copeland N.G, Gilbert D.J, Jenkins N.A, Berger R, Shaklai S, Amariglio N, Brok-Simoni F, Simon A.J, Rechavi G (2001). Nuclear membrane protein LAP2beta mediates transcriptional repression alone and together with its binding partner GCL (germ-cell-less). *J Cell Sci* **114**:3297-3307.
- Normanno D, Dahan M et Darzacq X (2012). Intra-nuclear mobility and target search mechanisms of transcription factors : a single-molecule perspective on gene expression. *BBA* **1819**:482-493.
- O'Neill L.P, Randall T.E, Lavender J, Spotswood H.T, Lee J.T et Turner B.M (2003). X-linked genes in female embryonic stem cells carry an epigenetic mark prior to the onset of X inactivation. *Hum Mol Genet* **12**:1783-1790.
- Ohlson R, Renkawitz R, Lobanenko V.V (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics, disease. *Trends Genet* **17**:520-527.
- Okabe I, Bailey LC, Attree O, Srinivasan S, Perkel JM, Laurent BC, Carlson M, Nelson DL, Nussbaum RL (1992). Cloning of human and bovine homologs of SNF2/SWI2: a global activator of transcription in yeast *S. cerevisiae*. *Nucleic Acids Res* **20**:4649-4655.
- Olson M.O.J, Dundr M, Szebeni A (2000). The nucleolus: an old factory with unexpected capabilities. *Trends cell biol* **10**:189-196.
- Ominchinski J.G, Clore G.M, Schaad O, Felsenfeld G, Trainor C, Appella E, Stahl S.J et Gronenborn A.M (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261**:438-446.
- Onate S.A, Tsai S.Y, Tsai M-J, et O'Malley B.W (1995). Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**:1354-1357.
- Ozer J, Moore P.A, Bolden A.H, Lee A, Rosen C.A et Lieberman P (1994). Molecular cloning of the small (γ) subunit of human TFIIA reveals functions critical for activated transcription. *Genes Dev* **8**:2324-2335.
- Pal M, McKean D et Luse D.S (2001). Promoter clearance by RNA polymerase II is an extended, multistep process strongly affected by sequence. *Mol Cell Biol* **21**:5815-5825.
- Pal M, Ponticelli A.S et Luse D.S (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell* **19**:101-110.
- Pant V, Kurukuti S, Pugacheva E, Shamsuddin S, Mariano P, Renkawitz R, Klenova E, Lobanenko V.V et Ohlson R (2004). Mutation in a single CTCF target site within the H19 imprinting control

- region lead to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol Cell Biol* **24**:3497-3504.
- Park J.T, Shih Ie M et Wang T.L (2008). Identification of Pbx1, a potential oncogene, as a Notch 3 target gene in ovarian cancer. *Cancer Res* **68**:8852-8860.
- Pascual-Garcia P, Govind C.K, Queralt E, Cuenca_Bono B, Llopis A, Chavez S, Hinnebusch A.G et Rodriguez-Navarro S (2008). Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. *Genes Dev* **22**:2811-2822.
- Paull T.T, Rogakou E.P, Yamazaki V, Kirchgessner C.U, Gellert M et Bonner W.M (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr Biol* **10**:886-895.
- Pedone P.V, Ominchinski J.G, Nony P, Trainor C, Gronenborn A.M, Clore G.M et Felsenfeld G (1997). The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. *EMBO J* **16**:2874-2882.
- Pehrson J.R et Fried V.A (1992). MacroH2A, a core histone containing a large non histone region. *Science* **257**:1398-1400.
- Perlmann T, Eriksson P et Wrange O (1990). Quantitative analysis of the glucocorticoid receptor-DNA interaction at the mouse mammary tumor virus glucocorticoid response element. *J Biol Chem* **265**:17222-17229.
- Perrini B, Piacentini L, Fanti L, Altieri F, Chichiarelli S, Berloco M, Turano C, Ferraro A et Pimpinelli S (2004). HPA controls telomere capping, telomere elongation, and telomere silencing by two different mechanisms in *Drosophila*. *Mol Cell* **15**:467-476.
- Peterson C.L, Dingwall A et Scott M.P (1994). Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. *Proc Natl Acad Sci USA* **91**:2905-2908.
- Peterson C.L et Laniel M.A (2004). Histones and histone modifications. *Curr Biol* **14**:R546-551.
- Pfahl M (1993). Nuclear receptor /AP-1 interaction. *Endocr Rev* **14**:651-658.
- Phillips J.E et Cores V.G (2009). CTCF: master weaver of the genome. *Cell* **137**:1194-1211.
- Polioudaki H, Kourmouli N, Drosou V, Bakou A, Theodoropoulos P.A, Singh P.B, Giannakouros T et Georgatos S.D (2001). Histones H3/H4 form a tight complex with the inner nuclear membrane protein LBR and heterochromatin HP1. *EMBO rep* **2**:920-925.
- Pombo A, Cuello P, Schul W, Yoon JB, Roeder R.G, Cook P.R, Murphy S (1998). Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J* **17**:1768-1778.
- Prasanth K.V et Spector D.L (2007). Eukaryotic regulatory RNAs: an answer to the «genome complexity» conundrum. *Genes Dev* **21**:11-42.

- Pratt W.B et Toft D.O (1997). Steroid receptor interaction with heat shock protein and immunophilin chaperones. *Endocr Rev* **18**:306-360.
- Quinn J, Fyrberg A.M, Ganster R.W, Schmidt M.C et Peterson C.L (1996). The yeast SWI/SNF complex has DNA-binding properties similar to HMG-box proteins. *Nature* **379**:844-847.
- Qumsiyeh M.B , 1999. Structure and function of the nucleus: anatomy and physiology of chromatin. *Cell Mol Life Sci* **55**:1129-1140.
- Rai K, Huggins I.J, James S.R, Karpf A.R, Jones D.A et Cairns B.R (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd 45. *Cell* **135**:1201-1212.
- Raisner R.M, Hartley P.D, Meneghini M.D, Bao M.Z, Liu C.L, Schreiber S.L, Rando O.J et Madhani H.D (2005). Histone variant H2A.Z marks the 5'ends of both active and inactive genes in euchromatin. *Cell* **123**:233-248.
- Raj D, Liu T, Samadashwily G, Li F et Grossman D (2008). Survivin repression by p53 Rb and E2F2 in normal human melanocytes. *Carcinogenesis* **29**:194-201.
- Ramakrishnan V, Finch J.T, Graziano V, Lee P.L et Sweet R.M (1993). Crystal structure of globular domain of histone H5 and its implication for nucleosome binding. *Nature* **362**:219-224.
- Rangasamy D, Berven L, Ridgway P et Tremethick D.J (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. *The EMBO J* **22**:1599-1607.
- Rayasam G.V, Elbi C, Walker D.A, wolford R.G, Fletcher T.M, Edwards D.P et Hager D.L (2005). Ligand specific dynamics of the progesterone receptor in living cells and during chromatin remodeling *in vitro*. *Mol Cell Biol* **25**:2406-2418.
- Razandi M, Pedram A, Rosen E et Levin E.R (2004). BRCA1 inhibits membrane estrogen and growth factor receptor signaling to cell proliferation in breast cancer. *Mol Cell Biol* **24**:5900-5913.
- Redon C, Pilch D, Rogakou E, Sedelnikova O, Newrock K et Bonner W.M (2002). Histone H2A variants H2AX and H2AZ. *Curr Opin Genet Dev* **12**:162-169.
- Reeves W.M et Hahn S (2003). Activator-independent functions of the yeast mediator sin4 complex in preinitiation complex formation and transcription reinitiation. *Mol Cell Biol* **23**:349-358.
- Reid G, Denger S, Kos M et Gannon F (2002). Human estrogen receptor- α : regulation by synthesis, modification and degradation. *Cell Mol Life Sci* **59**:821-831.
- Renaud J-P, Rochel N, Ruff M, Vivat V, Chambon P, Gronemeyer H et Moras D (1995). Crystal structure of the RAR- γ ligand-binding domain bound to all-trans retinoic acid. *Nature* **378**:681-689.
- Renauld H, Aparicio O.M, Zierath P.D, Billington B.L, Chhablani S.K et Gottschling D.E (1993). Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev* **7**:1133-1145.
- Riley T, Sontag E, Chen P et Levine A (2008). Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol* **9**:402-412.
- Rinn J.L et Huarte M (2011). To repress or not to repress: this is the guardian's question. *Trends Cell Biol* **21**:344-353.

- Rober R.A, Weber K et Osborn M (1989). Differential timing of nuclear lamin A/C expression in the various organs of the mouse embryo and the young animal: a developmental study. *Dev* **105**:365-378.
- Robyr D, Gégonne A, Wolffe A.P et Wahli W (2000). Determinants of vitellogenin B1 promoter architecture, HNF3 and estrogen responsive transcription within chromatin. *J Biol Chem* **275**:28291-28300.
- Rochette-Egly C (2005). Dynamic combinatorial networks in nuclear receptor-mediated transcription. *J Biol Chem* **280**:32565-32568.
- Rodriguez-Navarro S, Fischer T, Luo M.J, Antunez O, Brettschneider S, Lechner J, Perez-Ortin J.E, Reed R et Hurt E (2004). Sus 1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell* **116**:75-86.
- Rollins R.A, Morcillo P et Doresett D (1999). Nipped-B, a drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152**:577-593.
- Rosenfeld M.G, Lunyak U.V et Glass C.K (2006). Sensors and signals: a coactivators/corepressor/epigenetic code for integrating signal-dependant programs of transcriptional response. *Genes Dev* **20**:1405-1428.
- Rodriguez-Navarro S (2009). Insights into SAGA function during gene expression. *EMBO rep* **10**:843-850.
- Rubio E.D, Reiss D.J, Welsh P.L, Distèche C.M, Filipova G.N, Baliga N.S, Aebersold R, Ranish J.A et Krumm A (2008). CTCF physically links cohesin to chromatin. *PNAS* **105**:8309-8314.
- Samara N.L et Wolberger C (2011). A new chapter in the transcription SAGA. *Curr Opin Struc Biol* **21**:767-774.
- Samuelson C.O, Baraznenok V, Khorosjutina O, Spahr H, Kieselbach T, Holmberg S et Gustafsson C.M (2003). TRAP230/ARC240 and TRAP 240/ARC250 mediator subunits are functionally conserved through evolution. *Proc Natl Acad Sci USA* **100**:6422-6427.
- Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y et Hume D.A (2007). Mammalian RNA polymerase II core promoters : insights from genome-wide studies. *Nat Rev Genet* **8**:424-436.
- Sanyal A, Lajoie B.R, Jain G et Dekker K (2012). The long-range interaction landscape of gene promoters. *Nature* **489**:109-113.
- Schirmer E.C, Guan T, Gerace L (2001). Involvement of the lamin rod domain in heterotypic lamin interaction important for nuclear organization. *J Cell Biol* **153**:479-490.
- Schmid M, Arib G, Laemmli C, Nishikawa, Durussel T et Laemmli U.K (2006). Nup-PI: the nucleosome-promoter interaction of gene in yeast. *Mol Cell* **21**:379-391.
- Schmidt D, schwalie P.C, Ross-Innes C.S, Hurtado A, Brown G.D, Carroll J.S, Flicek P et Odom D.T (2010). A CTCF-independent role for cohesin in tissues-specific transcription. *Genome Res* **20**:578-588.

- Schultz D.C, Friedman J.R et Rauscher III F.J (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2 α subunit of NuRD. *Genes Dev* **15**:428-443.
- Schuur E.R, Loktev A.V, Sharma M, Sun Z, Roth R.A et Weigel R.J (2001). Ligand-dependent interaction of estrogen receptor- α with members of the forkhead transcription factor family. *J Biol Chem* **276**:33554-33560.
- Sem D.S, Casimiro D.R, Kliewer S.A, Provencal J, Evans R.M et Wright P.E (1997). NMR spectroscopic studies of the DNA-binding domain of the monomer-binding nuclear orphan receptor, human estrogen related receptor-2. The carboxyl-terminal extension to the zinc-finger region is unstructured in the free form of the protein. *J Biol Chem* **272**:18038-18043.
- Serandour A.A, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Metivier R, Salbert G et Eeckhoutte J (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**:555-565.
- Shandilya J et Roberts S.G (2012). The transcription cycle in eukaryotes: from peoductive initiation to RNA polymerase II recycling. *BBA* **1819**:391-400.
- Shang Y, Hu X, Di Renzo J, Lazar M.A et Brown M (2000). Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**:843-852. *J Cell Sci* **119**:4101-4116.
- Sharp Z.D, Mancini M.G, Hinojos C.A, Dai F, Berno V, Szafran A.T, Smith K.P, Lele T.T, Ingber D.E et Mancini M.A (2006). Oestrogen receptor alpha exchange and chromatin dynamics are ligand and domain-dependent. *J*
- Shen X, Mizuguchi G, Hamiche A et Wu C (2000). A chromatin remodeling complex involved in transcription and DNA processing. *Nature* **406**:541-544.
- Simpson E.R, Clyne C, Rubin G, Boon W.C, Robertson K, Britt K, Speed C et Jones M (2002). Aromatase--a brief overview. *Annu Rev Physiol* **64**:93-127.
- Sims R.J 3rd, Mandal S.S et Reinberg D (2004). Recent highlights of RNA-polymerase II mediated transcription. *Curr Opin Cell Biol* **16**:263-271.
- Singh B.N et Hampsey M (2007). A transcription-independent role for TFIIB in gene looping. *Mol Cell* **27**:806-816.
- Singh M; Popowicz G.M, Krajewski M et Holak T.A (2007). Structural ramification for acetyl-lysine recognition by the bromodomain of human BRG1 protein, a central ATPase of the SWI/SNF remodeling complex. *Chem Bio Chem* **8**:1308-1316.
- Smale S.T, Jain A, Kaufmann J, Emami K.H, Lo K et Garraway I.P (1998). The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. *Cold Spring Harb Symp Quant Biol* **63**:21-31.
- Smale S.T (2010). Pioneer factors in embryonic stem cells and differentiation. *Curr Opin Genet Dev* **20**:519-526.

- Somech R, Shaklai S, Geller O, Amariglio N, Simon A.J, Rechavi G et Gal-Yam E.N (2005). The nuclear-envelope protein and transcriptional repressor LAP2-beta interact with HDAC3 at the nuclear periphery, and induces histone H4 deacetylation. *J Cell Sci* **118**:4017-4025.
- Song C.X, Szulwach K.E, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C.H, Zhang W, Jian X, Wang J, Zhang L, Loonev T.J, Zhang B, Godley L.A, Hicks L.M, Lahn B.T, Jin P et He C (2011). Selective chemical labeling reveals the genome wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* **29**:68-72.
- Spann T, Moir R.D, Goldman A.E, Stick R et Goldman R.D (1997). Disruption of the nuclear lamin organization alters the distribution of replication factors and inhibits DNA synthesis. *J Cell Biol* **136**:1201-1212.
- Spann T, Goldman A.E, Wang C, Huang S et Goldman R.D (2002). Alteration of nuclear lamin organization inhibits RNA polymerase II-dependant transcription. *J Cell Biol* **156**:603-608.
- Sparmann A et van Lohuizen M (2006). Polycomb silencers control cell fate, development and cancer. *Nat Rev Cancer* **6**:846-856.
- Spector D.L (2003). The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem* **72**:573-608.
- Spilianakis C.G et Flavell R.A (2004). Long-range interchromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* **5**:1017-1027.
- Stavreva D.A, Varticovski L et Hager G.L (2012). Complex dynamics of transcription regulation. *BBA* **1819**:657-666.
- Stock J.K, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher A.G et Pombo A (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**:1428-1435.
- Strahl B.D et Allis C.D (2000). The language of covalent histone modifications. *Nature* **403**:41-45.
- Strohner R, Nemeth A, Jansa P, Hofmann-Rohrer U, Santoro R, Grummt G et Grummt I (2001). NoRc-a novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J* **20**:4892-4900.
- Strouboulis J and Wolffe A.P (1996). Functional compartmentalization of the nucleus. *J Cell Sci* **109**:1991-2000.
- Subramaniam N, Cairns W et Okret S (1998). Glucocorticoids repress transcription from a negative glucocorticoid response element recognized by two homeodomain-containing proteins, Pbx and Oct-1. *J Biol Chem* **273**:23567-23574.
- Sun J, Meyers M.J, Fink B.E, Rajendran R, Katznellenbogen J.A et Katznellenbogen B.S (1999). Novel ligands that function as selective estrogens or antiestrogens for estrogen receptor- α or estrogen receptor- β . *Endocrinol* **140**:800-804.
- Szulwach K.E, Li X, Li Y, Song C.H, Han J.W, Kim S, Namburi S, Hermetz K, Kim J.J, Rudd M.K, Yoon Y.S, Ren B, He C et Jin P (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of Human embryonic stem cells. *Plos Genet* **7**:e1002154.

- Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependant or independant of DNA synthesis. *Cell* **116**:51-61.
- Tahilianian M, Koh K.P, Shen Y, Pastor W.A, Bandukwala H, Brudno Y, Agarwal S, Iyer L.M, Liu D.R, Aravind L et Rao A (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**:930-935.
- Taube J.H, Allton K, Duncan S.A, Shen L et Barton M.C (2010). FOXA1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells. *J Biol Chem* **285**:16135-16144.
- Theodorou V, Stark R, Menon S et Carroll J.S (2013). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res* **23**-doi:10.1101/gr.139469.
- Tolhuis B, Palstra R.J, Splinter E, Grosveld F et de Laat W (2002). Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol Cell* **10**:1453-1465.
- Tremblay G.B, Tremblay A, Copeland N.G, Gilbert D.J, Jenkins N.A, Labrie F et Giguere V (1997). Cloning, chromosomal localization and functional analysis of the murine estrogen receptor β . *Mol Endocrinol* **11**:353-365.
- Tsai S.Y, Srinivasan G, Allan G.F, Thompson E.B, O'Malley B.W et Tsai M-J (1990). Recombinant human glucocorticoid receptor induces transcription of hormone response genes in vitro. *J Biol Chem* **265**:17055-17061.
- Tsai M-J et O'Malley B.W (1994). Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annu Res Biochem* **63**:451-486.
- Tsukiyama T et Wu C (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**:1011-1020.
- Tsukiyama T, Palmer J, Landel C.C, Shiloach J et Wu C (1999). Characterization of the imitation switch subfamily of ATP-dependant chromatin-remodeling factors in *saccharomyces cerevisiae*. *Genes Dev* **13**:686-697.
- Újvári A, Pal M et Luse D.S (2002). RNA polymerase II transcription complexes may become arrested if the nascent RNA is shortened to less than 50 nucleotides. *J Biol Chem* **277**:32527-32537.
- Umayahara Y, Kawamori R, Watada H, Imano E, Iwama N, Morishima T, Yamasaki Y, Kajimoto Y et Kamada T (1994). Estrogen regulation of the insulin-like growth factor I gene transcription involves an AP-1 enhancer. *J Biol Chem* **269**:16433-16442.
- Umesono K, Murakami K.K, Thompson C.C et Evans R.M (1991). Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell* **65**:1-20.
- Vaisanen S, Dunlop T.W, Sinkkonen L, Frank C and Carlberg C (2005). Spatio-temporal activation of chromatin in the human CYP24 gene promoter in the presence of 1 α ,25-Dihydroxyvitamin D3. *J Mol Biol* **350**:65-77.

- Vakoc C.R, Letting D.L, Gheldof N, Sawado T, Bender M.A, Groudine M, Weiss M.J, Dekker J et Blobel G.A (2005). Proximity among distant regulatory elements at the β -globin locus requires GATA-1 and FOG-1. *Mol Cell* **17**:453-462.
- Varga-Weisz P.D, Wilm M, Bonte E, Dumas K, Mann M et Becker P.B (1997). Chromatin remodeling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* **388**:598-602.
- Vegeto E, Allan G.F, Schrader W.T, Tsai M-J, McDonnell D.P et O'Malley B.W (1992). The mechanisms of RU486 antagonism is dependant on the conformation of the carboxy-terminal taim of the human progesterone receptor. *Cell* **69**:703-713.
- Vekhoff P, Caccaldi A, Polverari D, Pylouster J, Pisano C et Arimondo P.B (2008). Triplex formation on DNA targets : how to choose the oligonucleotide. *Biochemistry* **47**:12277-12289.
- Visser A.E, Eils R, Jauch A, Little G, Bakker P.J, Cremer T et Aten J.A (1998). Spatial distribution of early and late replicating chromatin in interphase chromosome territories. *Exp Cell Res* **243**:398-407.
- Voegel J.J, Heine M.J, Zechel C, Chambon P et Gronemeyer H (1996). TIF2, a 160kDa transcriptional mediator for the ligand-dependent activation function of AF-2 of nuclear receptors. *EMBO J* **15**:101-108.
- Volpi E.V, Chevret E, Jones T, Vatcheva R, Williamson J, Beck S, Campbell R.D, Goldsworthy M, Powis S.H, Ragoussis J et al (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J Cell Sci* **113**:1565-1576.
- von Mikecz A, Zhang S, Montminy M, Tan E.M, Hemmerich P (2000). CREB-binding protein (CBP)/p300 and RNA polymerase II colocalize in transcriptionally active domains in the nucleus. *J Cell Biol* **150**:265-273.
- Wagner R.L, Apriletti J.W, Mc Grath M.E, West B.L, Baxter J.D et Fletterick R.J (1995). A structural role for hormone in the thyroid hormone receptor. *Nature* **378**:690-697.
- Wan L.B, Pan H, Hannenhalli S, Cheng Y, Ma J, Fedoriw A, Lobanenko V.V, Latham K.E, Schultz R.M et Bartolomei M.S (2008). Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Dev* **135**:2729-2738.
- Wang Y, Yin L et Hillgartner F.B (2001). The homeodomain protein PBX and MEIS1 are accessory factors that enhance thyroid hormone regulation of the malic enzyme gene in hepatocytes. *J Biol Chem* **276**:23838-23848.
- Wang G, Balamotis M.A, Stevens J.L, Yamaguchi Y, Handa H et Berk A.J (2005). Mediator requirement for both recruitment and postrecruitment steps in transcription initiation. *Mol Cell* **17**:683-694.
- Wang Z, Zang C, Cui K, Schones D.E, Barski A, Peng W et Zhao K (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**:1019-1031.
- Wang B, Xiao Z, Ko H.L et Ren E.C (2010). The p53 response element and transcriptional repression. *Cell Cycle* **9**:870-879.

- Weber M, Davies J.J, Wittig D, Oakeley E.J, Haase M, Lam W.L et Schubeler D (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet* **37**:853-862.
- Wederell E.D, Bilenky M, Cullum R, Thiessen N, Dagpinar M, Delaney A, Varhol R, Zaho Y, Zeng T, Bernier B, Ingham M, Hirst M, Robertson G, Marra M.A, Jones S et Hoodless P.A (2008). Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res* **36**:4549-4564.
- Weigel D et Jackle H (1990). The Fork head domain: a novel DNA binding motif of eukaryotic transcription factors ? *Cell* **63**:455-456.
- Weis L et Reinberg D (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol Cell Biol* **17**:2973-2984.
- Wendt K.S, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahat K, Imamoto F, Aburatami H, Nakao M, Amamoto N, Maeshima K, Shirahige K et Peters J.M (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**:796-803.
- Wendt K.S et Peters J.M (2009). How cohesin and CTCF cooperate in regulating gene expression. *Chrom Res* **17**:201-214.
- Wehling M (1997). Specific, nongenomic actions of steroid hormones. *Annu Rev Physiol* **59**:365-393.
- Weigel N.L, Beck C.A, Estes P.A, Predergast P, Altmann M Christensen K et Edwards D.P (1992). Ligands induce conformational changes in the carboxyl-terminus of progesterone receptors which are detected by a site-directed-antipeptide monoclonal antibody. *Mom Endocrinol* **6**:1585-1597.
- West A.G et Fraser P (2005). Remote control of gene transcription. *Hum Mol Genet* **16**:213-220.
- Willy P.J, Umesono K, Ong E.S, Evans R.M, Heyman R.A et Mangelsdorf D.J (1995). LXR, a nuclear receptor that defines a distinct retinoid response pathway. *Genes Dev* **9**:1033-1045.
- Willy P.J et Mangelsdorf D.J (1997). Unique requirements for retinoic-dependent transcriptional activation by the orphan receptor LXR. *Genes Dev* **11**:289-298.
- Wilson B.J et Giguere V (2008). Meta-analyses of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer* **7**:49.
- Wilson T.E, Fahrner T.J, Johnson M et Milbrandt J (1991). Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science* **252**:1296-1300.
- Wilson T.E, Paulsen R.E, Padett K.A et Milbrandt J (1992). Participation of non-zinc finger residues in DNA binding by two nuclear orphan receptors. *Science* **256**:107-110.
- Wilson C.J, Chao D.M, Imbalzano A.N, Schnitzler G.R, Kingston R.E et Young R.A (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84**:235-244.
- Wolffe A.P (1998). Chromatin: structure and function. Academic Press: San Diego, California.

- Woodage T, Basrai M.A, Baxeavanis A.D, Hieter P et Collins F.S (1997). Characterization of the CHD family of proteins. *Proc Natl Acad Sci USA* **94**:11472-11477.
- Woodcock C.L (2006). Chromatin architecture. *Curr Opin Struct Biol* **16**:213-220.
- Wright W.E, Tesmer V.M, Huffman K.E, Levene S.D et Shay J.W (1997). Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev* **11**:2801-2809.
- Wu J.I, Lessard J et Crabtree G.R (2009). Understanding the words of chromatin regulation. *Cell* **136**:200-206.
- Wu H, D'Alessio A.C, Ito S, Wang Z, Cui K, Zhao K, Sun Y.E et Zhang Y (2011). Genome wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* **25**:679-684.
- Wyce A, Xiao T, Whelan K.A, Kosman C, Walter W, Eick D, Hughes T.R, Krogan N.J, Strahl B.D et Berger S.L (2007). H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within SAGA-related complex. *Mol Cell* **27**:275-288.
- Yamashita R, Suzuki Y, Wakaguri H, Tsuritani K, Nakai K et Sugano S (2006). DTBSS: DataBase of Human Transcription Star Sites, progress report 2006. *Nucleic Acids Res* **34**:D86-D89.
- Yao T-P, Ku G, Zhou N, Scully R et Livingston D.M (1996). The nuclear hormone receptor coactivator SRC-1 is a specific target of p300. *Proc Natl Acad Sci USA* **93**:10626-10631.
- Yeh H.Y, Cheng S.W, Lin Y.C, Yeh C.Y, Lin S.F et Soo V.W (2009). Identifying significant genetic regulatory networks in the prostate cancer from microarray data based on transcription factor analysis and conditional independency. *BMC Med Genomics* **2**:70.
- Yokomori K, Zeidler M.P, Chen J.L, Verrijzer C.P, Mlodzik M et Tijian R (1994). Drosophila TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes Dev* **8**:2313-2323.
- Yuan G.C, Liu Y.J, Dion M.F, Slack M.D, Wu L.F, Altschuler S.J et Rando O.J (2005). Genome-scale identification of nucleosome position in *S.cerevisiae*. *Science* **309**:626-630.
- Yusufzai T.M, Tagami H, Nakatani Y et Felsenfeld G (2004). CTCF tethers an insulators to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* **13**:291-298.
- Zaret K.S (1999). Developmental competence of the gut endoderm: genetic potentiation by GATA and HNF3/Fork head proteins. *Dev Biol* **209**:1-10.
- Zaret K.S (2002). Regulatory pahses of early liver develoment: paradigms of organogenesis. *Nat Rev Genet* **3**:499-512.
- Zaret K.S, Watts J, Xu J, Wandzioch E, Smale S.T et Sekiya T (2008). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors from the endoderm. *Cold Spring Harb Symp Quant Biol* **73**:119-126.
- Zechel C, Shen X.Q, Chambon P et Gronemeyer H (1994a). Dimerization interfaces formed between the DNA binding domains determine the cooperative binding of RXR/RAR and RXR/TR heterodimers to DR5 and DR4 elements. *EMBO J* **13**:1414-1424.

- Zechel C, Shen X.Q, Chen J.Y, Chen Z.P, Chambon P et Gronemeyer H (1994b). The dimerization interfaces formed between the DNA binding domains of RXR, RAR and TR determine the binding specificity and polarity of the full-length receptors to direct repeats. *EMBO J* **13**:1425-1433.
- Zemach A et Zilberman D (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol* **20**:R780-R785.
- Zhang H, Richardson D.O, Roberts D.N, Utley R, Erdjument-Bromage H, Tempst P, Cote J, Cairns B.R (2004). The Yaf9 component of the SWR1 and NuA4 complexes is required for proper gene expression, histone H4 acetylation , and Htz1 replacement near telomeres. *Mol Cell Biol* **24**:9424-9436.
- Zhang H, Roberts D.N et Cairns B.R (2005). Genome wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**:219-231.
- Zhang Y, Liang J, Li Y, Xuan C, Wang F, Wang D, Shi L, Zhang D et Shang Y (2010). CTCF acts upstream of FOXA1 and demarcates the genomic response to estrogen. *J Biol Chem* **285**:28604-28613.
- Zhao Y, Abreu E, Kim J, Stadler G, Eskiocak Y, Terns M.P, Terns R.M, Shay J.W et Wright W.E (2011). Processive and distributive extension of human telomere by telomerase under homeostatic and non equilibrium conditions. *Mol Cell* **42**:297-307.
- Zhong S, Salomoni P, Pandolfi P.P (2000). The transcriptional role of PML and the nuclear body. *Nat Cell Biol* **275**:20980-20984.
- Zhou W, Zhu P, Wang J, Pascual G, Ohgi K.A, Lozach J, Glass C.K et Rosenfeld M.G (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcripctional elongation. *Mol Cell* **29**:69-80.
- Zirbel R.M, Mathieu U.R, Kurz A, Cremer T et Lichter P (1993). Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boudaries. *Chromosome Res* **1**:93-106.

ANNEXE

VU :

Le Directeur de Thèse
(Nom et Prénom)

VU :

Le Responsable de l'École Doctorale

VU pour autorisation de soutenance

Rennes, le

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

VU après soutenance pour autorisation de publication :

Le Président de Jury,
(Nom et Prénom)

RESUME :

En réponse à son environnement composé de signaux endogènes et exogènes, une cellule doit pouvoir adapter son transcriptome et cela à travers une modulation fine de l'expression de ses gènes. Les mécanismes permettant une telle adaptation reposent sur de multiples paramètres, entre autre l'organisation du génome, que ce soit au niveau de sa séquence primaire ou de son organisation au sein de la chromatine qui est un support pour l'intégration de nombreuses informations (structurelles et épigénétiques). De plus, l'organisation tridimensionnelle du noyau cellulaire apporte des contraintes physiques et fonctionnelles qui contribuent également à ces régulations. Afin de comprendre comment toutes ces informations peuvent être intégrées lorsqu'un signal régule la transcription d'un ensemble de gènes colinéaires ("cluster" de gènes), nos études se sont focalisées sur la description et dissection des mécanismes impliqués dans la régulation coordonnée de gènes œstrogéno-dépendants par le récepteur aux œstrogènes (ER) dans des cellules cancéreuses d'origine mammaire. Dans ce cadre, nous nous sommes plus particulièrement intéressés au cluster TFF, situé sur le bras long du chromosome 21, incluant le gène modèle *TFF1*, en utilisant des techniques d'analyse à grande échelle (ChIP-chip, ChIP-seq, 4C et analyses transcriptomiques).

La première partie de ce travail a consisté en (i) l'étude de ces mécanismes en intégrant les aspects régulateurs à la fois dans l'espace et le temps; ainsi qu'en (ii) l'identification des facteurs protéiques collaborant avec ER pour la mise en place de ces processus. Cette étude, menée dans deux types cellulaires différents (les MCF-7 qui expriment ER de façon endogène et MDA::ER qui l'expriment de façon ectopique) a permis de mettre en évidence un réseau d'interactions fonctionnelles cellule-spécifiques mettant en jeu des contacts dynamiques entre sites de liaison d'ER distants et promoteurs de ces gènes cibles, ainsi qu'un rôle des protéines de la famille GATA dans l'établissement et/ou la fonctionnalité de ce réseau.

La seconde partie de ce travail de thèse a porté plus particulièrement sur une autre famille de protéines impliquées dans l'établissement du cistrome ER: les facteurs FOX. En effet, la ré-expression de ER dans les cellules MDA::ER ne reproduit ni un cistrome ER ni un transcriptome œstrogéno-dépendant ressemblant à ceux des cellules MCF-7. Nous avons posé l'hypothèse que ceci pouvait être dû à l'absence dans les MDA::ER du facteur pionnier FOXA1. Après la création d'une lignée cellulaire MDA::ER exprimant FOXA1 et les analyses subséquentes, nous avons pu caractériser des mécanismes d'interrelations interférentes entre le FOXA1 ectopique et du FOXA2 endogène qui, pourraient être à l'origine du fait que FOXA1 soit incapable de réorienter les cellules MDA::ER vers un phénotype prolifératif œstrogéno-dépendant ressemblant à celui des cellules MCF-7.